

AFGL-TR-88-0327

INFERRING RELATIVE HUMIDITY PROFILES FROM 3DNEPH CLOUD DATA:
TECHNIQUE DEVELOPMENT AND DATA IMPACT STUDY

Thomas Nehrkorn and Ross N. Hoffman

Atmospheric and Environmental Research, Inc.
840 Memorial Drive
Cambridge, Massachusetts 02139

December 13, 1988

DTIC
ELECTE
MAR 29 1989
S D

Scientific Report No. 5

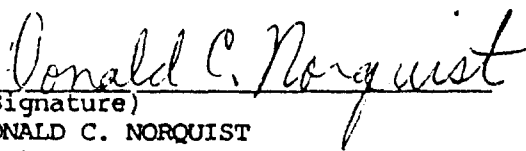
Approved for public release; distribution unlimited

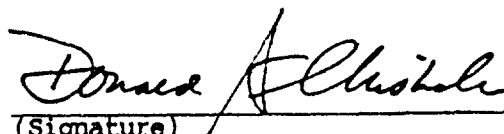
AIR FORCE GEOPHYSICS LABORATORY
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
HANSCOM AFB, MASSACHUSETTS 01731-5000

AD-A205 825

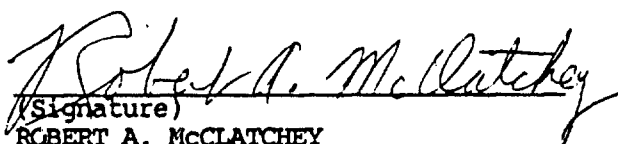
89 3 28 081

"This technical report has been reviewed and is approved for publication"


(Signature)
DONALD C. NORQUIST
Contract Manager


(Signature)
DONALD A. CHISHOLM
Branch Chief

FOR THE COMMANDER


(Signature)
ROBERT A. MCCLATCHEY
Division Director

This report has been reviewed by the ESD Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS).

Qualified requestors may obtain additional copies from the Defense Technical Information Center. All others should apply to the National Technical Information Service.

If your address has changed, or if you wish to be removed from the mailing list, or if the addressee is no longer employed by your organization, please notify AFGL/DAA, Hanscom AFB MA 01731-5000. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE					
4. PERFORMING ORGANIZATION REPORT NUMBER(S)			5. MONITORING ORGANIZATION REPORT NUMBER(S) AFGL-TR-88-0327		
6a. NAME OF PERFORMING ORGANIZATION Atmospheric and Environmental Research, Inc.		6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Air Force Geophysics Laboratory		
6c. ADDRESS (City, State, and ZIP Code) 840 Memorial Drive Cambridge, MA 02139			7b. ADDRESS (City, State, and ZIP Code) Hanscom AFB, MA 01731-5000		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Air Force Geophysics Laboratory		8b. OFFICE SYMBOL (If applicable) AFGL/LY	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F19628-86-C-0141		
8c. ADDRESS (City, State, and ZIP Code) Hanscom AFB, MA 01731-5000			10. SOURCE OF FUNDING NUMBERS		
			PROGRAM ELEMENT NO. 61102F	PROJECT NO. 6670	TASK NO. 10
					WORK UNIT ACCESSION NO. CB
11. TITLE (Include Security Classification) Inferring Relative Humidity Profiles from 3DNEPH Cloud Data: Technique Development and Data Impact Study					
12. PERSONAL AUTHOR(S) T. Nehr Korn and R.N. Hoffman					
13a. TYPE OF REPORT Scientific Report #5		13b. TIME COVERED FROM TO		14. DATE OF REPORT (Year, Month, Day) 1988 December 13	
				15. PAGE COUNT 50	
16. SUPPLEMENTARY NOTATION <i>Supplementary notes</i>					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB-GROUP	Cloud Data, Humidity Data, Data Assimilation, Numerical Weather Prediction, Observing System Experiment.		
19. ABSTRACT (Continue on reverse if necessary and identify by block number) The inference of profiles of relative humidity from cloud data was investigated in a co- location study of 3DNEPH and radiosonde data over North America. Regression equations were developed for the first two EOFs of relative humidity, using vertically compacted and horizontally averaged 3DNEPH cloud cover values as predictors. The regression equa- tions were found to have smaller errors than existing level-to-level cloud to humidity conversion techniques in tests in which no attempt was made to tune the existing methods for optimal performance. The utility of the bogus RH data for operational data assimila- tion was studied in a global observing system experiment, using the AFGL global data assimilation system. The impact of the additional RH data, which was supplied to the moisture analysis in the 30°-50°N latitude belt, was evident in analyses in the early part of the assimilation runs, but was lost in the noise of the assimilation system at later times. There was no measurable impact on the quality of the RH analyses or fore- casts. Reasons for this lack of impact are related to deficiencies in the assimilation system and the large observation errors of the bogus RH data.					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input type="checkbox"/> UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified		
22a. NAME OF RESPONSIBLE INDIVIDUAL D. Norquist			22b. TELEPHONE (Include Area Code) 617-377-2962		22c. OFFICE SYMBOL AFGL/LYP

1. Introduction

The analysis of atmospheric moisture remains a challenging problem in data assimilation today. Relative humidity (RH) varies on horizontal scales that are smaller than the typical separation distance of radiosondes over Northern Hemisphere continents, and there are virtually no radiosonde measurements over large portions of the Southern Hemisphere and the Northern Hemisphere oceans. Satellite observations may help to improve our knowledge of the RH field.

Moisture information may be obtained from satellite sounding data by retrieval techniques similar to those used for temperature retrievals, or by use of cloudiness information contained in satellite imagery. Since moisture (and temperature) retrieval, at least from infrared radiances, is difficult in cloudy atmospheres, there is a possibility that inference of humidity data from cloudiness information may successfully supplement moisture retrieval from radiance data. Here we study the inference of humidity profiles from cloud data, using the 3DNEPH data base as the source of the cloudiness information. The 3DNEPH (now RTNEPH) is a high resolution cloud data base produced operationally by the US Air Force Global Weather Central (AFGWC). A collocation study of cloud data with radiosonde measurements of relative humidity is used to develop and test a statistical method for inferring humidity profiles; a global data impact study is used to assess the utility of this moisture information. In this report we review some of the previously developed methods for inferring humidity from cloud cover data, describe the data base and processing used in our collocation study, and discuss the development and testing of the new method for inferring humidity. We then describe the data impact test and summarize our results and conclusions.

2. Background

Several different techniques for inferring relative humidity from cloud information exist. They may be grouped into two categories: level-by-level approaches, which use relationships between relative humidity and cloud cover at a particular level or layer of the atmosphere, and profile approaches, which infer vertical profiles of relative humidity from cloud information.

Of the level-by-level approaches, the one most widely used is described in Chu and Parrish (1977). As implemented by Tibaldi (1982), humidity is determined in the boundary layer (assumed to be 50 hPa thick) and three layers in the troposphere between the boundary layer and 300 hPa from cloudiness observations (both from surface observers and satellite observations), using a relationship of the form:

$$RH = M - A \cos (\pi \cdot \eta), \quad (1)$$

where RH is the relative humidity, η the fractional cloud cover, and the coefficients M and A depend on the humidity layer (see Appendix A).

The AFGWC uses humidity - cloudiness relationships within its 3DNEPH (now RTNEPH) analysis program. As described in Fye (1978), relative humidity is expressed in terms of a condensation pressure spread (CPS), which is the pressure increment an air parcel needs to be lifted to reach saturation. The condensation spread is then related to cloud cover by an empirically derived curve for each mandatory pressure level (see Appendix A).

Rasmussen (1982) used 3DNEPH data to derive multiple regression equations relating mandatory level relative humidity to cloud information. Separate regression equations were used for the different mandatory levels.

Relationships between relative humidity and cloud cover are also used within the interactive radiation parameterizations of some prediction models. For example, Geleyn (1981) describes one such relationship as

$$\eta = \max\left[0, \frac{RH - RH_c}{1 - RH_c}\right]^2 \quad (2)$$

where the critical relative humidity RH_c depends on pressure (see Appendix A). The above relationship is easily inverted to obtain bogus humidity from cloudiness.

Saito and Baba (1988) investigated level-by-level approaches in a collocation study over the western Pacific Ocean. The cloud cover data was derived from infrared imagery observed by the GMS satellite. They arrived at a modified form of equation (2), in which a small but nonzero cloud cover was allowed even for subcritical relative humidities.

Norquist (1988) investigated several of these level-by-level approaches in a global colocation study using 3DNEPH data. Based on a comparison of the Tibaldi, AFGWC, and inverse ECMWF schemes, he found the Tibaldi method to have the smallest errors. He demonstrated the potential usefulness of bogus RH data inferred from 3DNEPH cloud cover data with data assimilation experiments in which the bogus RH was used to replace RH data measured by radiosondes.

A potential problem with all of these level-by-level approaches is the rather large uncertainty of the cloud height assignment in most cloud cover data, which will lead to errors in cloud cover - relative humidity relationships. In addition, cloud cover at a given level is related to moisture (and other atmospheric variables) at other levels in some meteorological situations, most notably for the case of convective clouds. This problem is addressed by the profile approaches, which attempt to retrieve an entire vertical profile of moisture from cloud cover data.

An example of the profile approach is the technique used at the Japanese Meteorological Service and described in Kanamitsu (1984). Cloud information including cloud cover, variability of cloud cover and cloud top are used to identify one of 60 different categories, each of which is associated with a typical relative humidity profile. The categories are defined a priori.

A similar approach was used by Mills and Davidson (personal communication, 1987), who used total cloud cover, the variance of cloud top temperature, and the height of maximum cloudiness to define categories with typical RH profiles.

The aim of our regression study is to develop statistical methods to infer RH profiles from the 3DNEPH cloud data base. We used the empirical orthogonal functions (EOFs) of RH to determine the important features of the observed RH profiles. The EOF coefficients were related to colocated cloud data by means of multivariate linear regression equations. The data used in the regression study was restricted to the North American continent, resulting in a homogeneous sample of high quality radiosonde measurements. The data base is described in more detail in the next section.

Accession For	
NTIS CRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and for
A-1	several

3. Data Base and Processing

The data base used in the colocation study consists of radiosonde and cloudiness data over North America during February and June of 1979.

The radiosonde data were extracted from the final, reprocessed FGGE II b data set, and subjected to an additional quality control. Both mandatory and significant level data were used. A total of 2699 (2235) soundings, for 97 stations were extracted from the FGGE data set for 00 UTC, 5 February through 00 UTC, 22 February (00 UTC, 14 June through 00 UTC, 28 June). The location of the radiosonde stations are shown in Fig. 1. A preliminary investigation using mandatory pressure levels revealed a large number of missing values at 1000 hPa and 850 hPa, and a dependence of the RH profile, when defined with respect to pressure levels, on station elevation. For these reasons, the relative humidity data were interpolated from the pressure levels to a terrain-following coordinate system, and an EOF analysis was performed on the result. The coordinate system was a modified sigma coordinate:

$$\sigma = (p - p_t) / (p_s - p_t), \text{ where } p_t = 300 \text{ hPa and } p_s = \text{surface pressure.}$$

A total of 15 equally spaced levels were used between $\sigma=1$ and $\sigma=0.15$. Only soundings with enough measurements to allow this interpolation were used in the EOF computations; this reduced the total number of soundings to 1918 (1724) for February (June). Given the EOFs, any relative humidity profile (RH_k , $k = 1, 15$) may be expressed in terms of an EOF expansion as follows

$$RH_k = \overline{RH}_k + \sum_{m=1}^{15} e_m E_{m,k},$$

where \overline{RH}_k is the climatological mean RH profile, e_m the coefficient of the m^{th} EOF, and $E_{m,k}$ the value of the m^{th} EOF at level k . We computed the EOFs as the eigenvectors of the relative humidity covariance matrix. Two separate sets of EOFs were computed for each month, one (denoted EOF-All) based on the covariance matrix about the mean of all radiosondes in the extracted data set for each month, another (denoted EOF-Bands) based on the covariance matrix

about the mean humidity profiles computed separately for three latitude bands of width 10° , spanning 20°N to 50°N . The mean profiles and associated standard deviations are shown in Fig. 2 for EOF-All, and Fig. 3 for EOF-Bands. Except for the February means, differences between the band statistics are small. The profiles of the first three EOFs are shown in Fig. 4 and 5 for EOF-All and EOF-Bands, respectively. The leading EOFs are similar for February and June; EOF-All and EOF-Bands have basically the same structure. The amount of variance explained by these EOFs is shown in Table 1. In comparing numbers for EOF-All and EOF-Bands one must bear in mind that the total amount of variance about the mean is somewhat smaller for EOF-Bands than EOF-All.

Table 1: Percent of variance explained by the RH EOFs. EOF-All and EOF-Bands refer to the EOFs defined with respect to latitude-independent and latitude-dependent mean RH profiles, respectively.

EOF No.	February		June	
	EOF-All	EOF-Bands	EOF-All	EOF-Bands
1	54.4	51.5	44.6	45.1
2	22.6	23.4	21.6	21.1
3	8.0	8.7	10.4	10.5
4	4.9	5.3	7.0	6.7
5	2.9	3.2	4.4	4.5
6	1.9	2.1	2.8	2.9
7	1.3	1.4	2.2	2.3
8	1.0	1.1	1.7	1.7
9	0.8	0.8	1.2	1.2
10	0.6	0.6	1.0	1.0
11	0.4	0.5	0.8	0.8
12	0.4	0.4	0.7	0.7
13	0.3	0.4	0.6	0.6
14	0.3	0.4	0.5	0.5
15	0.3	0.3	0.3	0.3

The cloudiness data base used for this study is the Air Force 3DNEPH analysis. The 3DNEPH data set is a global gridded data set with a resolution of 47.6 km on a polar stereographic grid (the so-called 8th-mesh grid). The grid for each hemisphere is subdivided into 64 boxes, each of which contains 64x64 gridpoints. The data at each gridpoint consist of percent cloud cover for total sky cover and for 15 layers in the vertical, as well as several other parameters (cloud type, base and top heights, terrain height, and pres-

ent weather; see Appendix B for a detailed description of the parameters and their recoding for this colocation study). In addition, a vertically compacted set of cloud cover values, which correspond to boundary layer clouds and cloud cover for layers surrounding the 6 mandatory levels between 1000 hPa and 300 hPa were derived from the 15 layer values. The vertical compaction, described in more detail in Appendix B, reduces the data volume without a significant loss of information since the cloud cover values in the 15 3DNEPH layers are partially redundant. The 3DNEPH data is derived primarily from IR and visible satellite imagery, and supplemented with conventional surface, radiosonde, and aircraft reports. For the present study we extracted 3DNEPH data for the gridpoints closest to the extracted radiosonde stations; in addition to the values at the gridpoint itself (referred to in the following as central values), the neighboring 24 points were used to compute means and standard deviations. Data were extracted for the 3DNEPH boxes 43, 44, and 45. The 69 sounding locations that fall within the extracted 3DNEPH boxes are shown in Fig. 1 as stars.

4. Regression Study

The data base described in the previous section was used to develop and test a new method to infer a RH profile from 3DNEPH data. The data set for each month was subdivided into a dependent and independent sample; the former was used to develop multiple regression equations for the EOF coefficients of relative humidity, the latter was used to validate the regression equations and to compute error statistics of the bogus RH data.

The dependent sample for February, which covers data from 00 UTC 5 February 1979 to 12 UTC 16 February 1979, was used for some exploratory, univariate correlation calculations between 3DNEPH data and EOF coefficients. Aside from identifying the most promising-looking predictors from the 3DNEPH data set, these computations were used to assess the effect of stratifying the sample into various subsets. One such stratification, based on the time of day, showed correlation coefficients to be consistently higher for the 00 UTC sample than for the 12 UTC, or the aggregate 00 UTC and 12 UTC sample. Control runs using a random subsample showed these differences to be significant. The most likely reason is that the cloud data over the U.S. are more reliable and less noisy at 00 UTC, because the primary data source of the

3DNEPH is IR imagery and in the late afternoon the temperature contrast between the land surface and the cloud tops is greatest.

Correlation coefficients between one of the 3DNEPH variables and the coefficients of either EOF 1 or 2 showed no consistent differences between latitude bands (30° - 40° N and 40° - 50° N); similar correlation coefficients for EOF 3 were nonzero only in the 30° - 40° N sample. Finally, these correlation coefficients of EOF 1 and 2 were consistently higher for the horizontally averaged 3DNEPH values than the corresponding central values, while the standard deviations of the 3DNEPH variables showed no useful correlations at all.

A stepwise regression procedure was used to derive multiple regression equations for the coefficients of EOF 1 and 2. Stepwise regression is a method for determining the "best" set of predictors (Neter and Wasserman, 1974, Chapter 11; Draper and Smith, 1966, Chapter 6). In our case, some of the 3DNEPH data have large biases or random observational errors and many of the variables are intercorrelated. We have the original data, the vertically compacted data, as well as the local area average and variance of the original and vertically compacted data. On the dependent sample, we are sure to explain more of the variance of the EOF coefficients for each predictor we add to our regression equations. However, using all potential predictors is sure to overfit the data, resulting in poorer performance on the independent sample. Therefore we must employ some means of deciding which predictors to use.

The stepwise regression procedure iteratively adds and deletes variables from the prediction equation. At each step, all possible predictors not yet included in the regression are considered for inclusion. The one reducing the unexplained variance the most is added to the regression if this reduction in variance is significant, as judged by the coefficient of partial correlation, which is a type of F statistic (op. cit.). Since the predictor added may be highly correlated with a previous predictor, all current predictors are considered for deletion. The predictor having the smallest partial correlation is deleted if the associated reduction in variance is so small as to be insignificant. The iteration stops when no further changes are made. The algorithm we used is based on that of Efroymson (1960). Based on the exploratory, univariate correlation statistics, some preliminary choices of predictors were made: In order to use the best available data for the development of the mul-

multiple regression equations, only 00 UTC data was used; multiple regression equations were based on the horizontal mean values only (tests of multiple regression equations using central values resulted in less skillful predictions of the EOF coefficients); multiple regression equations for EOFs 1 and 2 were developed without any stratification based on latitude. The multiple regression equation for EOF 3 for the 30°-40°N latitude band was able to explain only 30% of its variance. Since EOF 3 only contributes a small amount of the RH variance, its prediction was not pursued further. Two separate regression equations were developed for each EOF, one using only the vertically compacted cloud cover data, and another using all mean 3DNEPH variables. In the latter case, cloud cover data for 3DNEPH layers were added to the predictor sets by the stepwise regression algorithm only after the vertically compacted predictors were exhausted. The corresponding increases in the amount of the explained variance were rather modest, however, because of high correlations between the vertically compacted and the layer data. An example of this is shown in Fig. 6: scatterplots of EOF 1 coefficients versus cloud cover in layer 10 show a relatively high correlation, whereas the residuals from the regression equation using the vertically compacted data are already much less correlated, and including layer 10 cloud cover in the regression then removes the remaining correlation. Table 2 shows the predictor sets and the fraction of the explained variance (r^2) for the coefficients of EOF 1 and 2, both for the global mean and latitude-dependent mean (EOF-All and EOF-Bands), for February and June. We note that the r^2 values are slightly lower for EOF-Bands than EOF-All.

An evaluation of these regression equations was performed both for the dependent sample and the independent sample. For a comparison of the regression equation and existing methods of estimating humidity the bogus RH profiles were interpolated from the σ -coordinate system to mandatory pressure levels. For the purposes of interpolating, RH was assumed to vary linearly as a function of the logarithm of pressure. The interpolated bogus RH profiles were then compared with colocated RAOB data, along with bogus RH obtained by existing methods. Bias and rmse statistics were computed over all colocated 3DNEPH data points, even if the cloud cover used in the inference was zero; to reduce the positive bias likely to result from this procedure, the bogus humidity predicted by the existing methods was modified to be no higher than climatology in cases of zero cloud cover. A recalculation of the statistics,

Table 2: Predictor sets and r^2 for the regression equations. Results are shown separately for EOF 1 and 2. The predictor set of the vertically compacted set is shown first, additional 3DNEPH data of the second regression equation are shown in the second column; η denotes cloud cover, r^2 the fraction of variance explained by the regression.

Regression equation	<u>Vertically compacted data</u>		<u>Full 3DNEPH data set</u>	
	Predictors	r^2	Predictors	r^2
<u>EOF 1:</u>				
EOF-All February	η at 850, 700, 500 hPa	.554	η at layer 10	.568
EOF-Bands February	η at 700, 500, 400 hPa	.514	η at layer 10	.530
EOF-All June	η at 700, 400 hPa	.610	η at layer 15	.623
EOF-Bands June	η at 700, 400 hPa	.606	η at layer 15	.618
<u>EOF 2:</u>				
EOF-All February	η at 850, 700 500, 300 hPa	.413	η at layer 9	.434
EOF-Bands February	η at 850, 700 500, 300 hPa	.394	η at layer 9	.414
EOF-All June	terrain height η at 850, 500 hPa	.333	η at 700 hPa layer 8, 12 low cloud type	.484
EOF-Bands June	terrain height η at 700, 400 hPa	.328	η at layer 8,12 low, middle cloud type	.481

excluding all cases in which cloud cover values used as predictors were less than 10% showed that the results were not significantly affected by the use of cases with zero cloud cover.

Table 3 shows the RMS error and bias of three existing methods for the time period corresponding to the dependent sample in February. The smaller

sample sizes at the 1000 hPa and 850 hPa levels shown in Table 3 are due to observations with small surface pressures, i.e. high station elevation. As an independent reference, errors associated with climatology are shown as well; climatology is defined here as the average over all radiosonde observations in the extracted set within the 3DNEPH boxes, computed separately for each month. Errors for the GWC and Tibaldi method are roughly comparable; both methods show some skill when compared to climatology. The inverse ECMWF formula suffers from a rather large bias. The corresponding statistics for the bogus RH profiles based on the regression equations for EOFs 1 and 2 are shown in Table 4. The smaller sample sizes at the 400 hPa level shown in Table 4 are caused by observations with surface pressure of above 967 hPa, for which the top sigma level is below 400 hPa. It can be seen that the regression equations result in smaller errors than the existing methods; differences between the different regression equations, i.e. those for EOF-All or EOF-Bands, and those using the vertically compacted or the full 3DNEPH data set are generally small.

Table 3: Bias and RMS errors for the existing methods, for the dependent sample in February. The sample sizes for each statistic are shown in parentheses. The row labeled "Average" is an unweighted average of the level values.

	AFGWC		Tibaldi		inverse ECMWF		Climatology	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1000 hPa	12.01	17.87	12.03	17.71	35.32	40.96	7.36	21.49
	(82)		(82)		(82)		(211)	
850 hPa	8.85	26.09	12.45	27.69	19.69	31.02	5.71	32.45
	(276)		(276)		(276)		(440)	
700 hPa	14.44	31.50	12.30	30.18	25.08	37.31	2.47	33.28
	(302)		(302)		(302)		(470)	
500 hPa	1.21	26.13	3.92	26.21	23.35	34.48	- .77	31.61
	(303)		(303)		(303)		(468)	
400 hPa	-5.33	23.24	4.25	23.48	28.95	37.33	-2.50	28.89
	(303)		(303)		(303)		(468)	
Average	6.24	24.97	8.99	25.05	26.48	36.22	2.45	29.54

Table 4: As Table 3, but for the regression equations. Set 1 and 2 refer to the vertically compacted and the full 3DNEPH data set, respectively; EOF-All and EOF-Bands denote EOFs based on latitude-independent and latitude-dependent mean RH profiles, respectively.

	EOF-All, set 1		EOF-All, set 2		EOF-Bands, set 1		EOF-Bands, set 2	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1000 hPa	10.83	18.80	10.53	18.72	9.37	18.07	9.14	18.12
	(82)		(82)		(82)		(82)	
850 hPa	2.13	23.27	1.99	23.07	1.43	22.46	1.36	22.38
	(276)		(276)		(276)		(276)	
700 hPa	-.77	26.58	-1.18	26.36	-1.00	26.28	-1.21	25.82
	(289)		(289)		(289)		(289)	
500 hPa	1.31	25.44	.69	25.28	1.03	25.06	.63	24.96
	(291)		(291)		(291)		(291)	
400 hPa	4.20	22.40	2.78	22.88	6.17	21.94	5.18	22.44
	(106)		(106)		(106)		(106)	
Average	3.54	23.30	2.96	23.26	3.40	22.76	3.02	22.75

Table 5 and Table 6 show the verification statistics for the independent sample in February, which covers the period from 00 UTC, 17 February, through 00 UTC, 22 February. Because of the small differences between the regression equations using the vertically compacted and the full 3DNEPH data set, only the results for the former are shown here. The results are qualitatively the same for the two samples in February. As is to be expected, the errors for the regression equations are somewhat larger for the independent than the dependent sample, but they are larger for the existing methods, as well.

Table 5: As Table 3, except for the independent sample in February.

	AFGWC		Tibaldi		inverse ECMWF		Climatology	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1000 hPa	2.29	17.78	2.65	17.20	26.10	32.72	-7.74	20.80
	(104)		(104)		(104)		(230)	
850 hPa	12.76	29.03	15.65	27.27	23.10	35.03	1.56	33.51
	(254)		(254)		(254)		(412)	
700 hPa	20.59	34.14	16.94	32.33	30.97	41.38	4.44	33.59
	(296)		(296)		(296)		(457)	
500 hPa	-7.90	27.55	11.17	29.26	29.84	39.52	4.65	32.06
	(288)		(288)		(288)		(458)	
400 hPa	-3.46	27.63	7.58	27.50	30.49	41.49	-.23	29.96
	(298)		(298)		(298)		(469)	
Average	8.14	27.22	10.80	27.59	28.10	38.03	0.54	29.98

Table 6: As Table 4, except for the independent sample in February. Results are shown for the regression equations based on the vertically compacted 3DNEPH data set only.

	EOF-All		EOF-Bands	
	Bias	RMSE	Bias	RMSE
1000 hPa	-1.01	17.81	-1.75	18.35
	(104)		(104)	
850 hPa	3.81	25.39	3.79	26.44
	(254)		(254)	
700 hPa	5.36	26.60	5.01	27.31
	(284)		(284)	
500 hPa	7.86	25.87	7.21	25.33
	(276)		(276)	
400 hPa	5.26	24.93	4.39	23.76
	(71)		(71)	
Average	4.26	24.12	3.73	24.24

The error statistics for the existing methods computed here can be compared to the results obtained by Norquist (1988) in his global colocation study for February of 1979. His results are similar in that the errors of the

AFGWC and Tibaldi methods are roughly comparable, with slightly smaller errors for the Tibaldi method. In general, his results show somewhat smaller errors for all three methods, except for the AFGWC method at 400 hPa, which shows smaller errors in our results. The positive bias of the inverse ECMWF method found here is consistent with his results, and also consistent with the results of Saito and Baba (1988).

Tables 7 and 8 contain the verification statistics for the independent sample in June, which covers the period 12 UTC, June 17, through 00 UTC, June 24. They show a smaller variance about climatology than for February, in agreement with the results shown in Fig. 2. The errors for the existing methods, on the other hand, show little change compared to February, resulting in only marginal or no skill over climatology for the AFGWC or the Tibaldi method. The errors of the regression equations are smaller for June than for February, and they are consistently smaller than those of climatology or the existing methods.

Table 7: As Table 5, but for the independent sample in June.

	AFGWC		Tibaldi		inverse ECMWF		Climatology	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1000 hPa	6.82 (118)	16.95	2.72 (118)	15.34	27.68 (118)	33.35	-3.31 (298)	17.62
850 hPa	8.16 (540)	21.47	7.50 (540)	21.70	17.15 (540)	25.59	-2.06 (774)	23.38
700 hPa	20.20 (585)	30.90	15.74 (585)	28.29	30.06 (585)	38.10	4.11 (809)	27.31
500 hPa	14.77 (564)	27.44	19.99 (564)	31.21	34.80 (564)	42.09	14.59 (784)	30.26
400 hPa	10.84 (584)	25.09	22.76 (584)	31.90	39.03 (584)	46.45	16.88 (804)	28.89
Average	12.16	24.37	13.74	25.69	29.74	37.12	6.04	25.49

Table 8: As Table 6, but for the independent sample in June.

	EOF-All		EOF-Bands	
	Bias	RMSE	Bias	RMSE
1000 hPa	-2.57 (118)	16.13	-4.06 (118)	16.04
850 hPa	-6.51 (540)	20.88	-6.30 (540)	20.80
700 hPa	-5.54 (557)	26.06	-4.79 (557)	25.88
500 hPa	3.95 (539)	23.60	3.52 (539)	23.55
400 hPa	7.82 (196)	22.65	8.06 (196)	22.63
Average	-.57	21.87	-.71	21.78

In summary, we identified regression equations for the first two EOFs of relative humidity; several different possibilities were investigated which differed in the definition of the RH EOFs and in the subset of the 3DNEPH data considered as predictors. Based on tests using the dependent and independent samples, it was found sufficient to use regression equations for the coefficients of EOF-All, i.e. the EOFs based on the mean RH profile computed for all latitudes between 20° and 50°N, and to use the vertically compacted, horizontally averaged 3DNEPH data set. Comparison with existing level-to-level methods for inferring RH from cloud cover showed the regression equations to perform better for both the dependent and independent samples. It should be noted that the verification was performed over the same geographical region for which the regression equations were developed, and that no attempt was made to tune the existing methods to improve their performance. The comparison with the level-to-level methods was performed primarily to provide some independent reference point for the errors associated with the regression equations, and to demonstrate the feasibility and potential utility of the approach.

5. Data Impact Study

The method for inferring relative humidity from 3DNEPH data described in the previous section was used in a data impact study using the global data assimilation system (GDAS) of the Air Force Geophysics Laboratory (AFGL). The AFGL GDAS consists of three major components: a global spectral forecast model

(GSM), an optimum interpolation analysis (OI), and a nonlinear normal mode initialization (NMI). The global spectral model is based on the NMC GSM designed by Sela (1980); the physics routines were taken almost intact from NMC (circa 1983), whereas the hydrodynamics were completely redesigned (Brenner et al., 1982, 1984). The optimum interpolation analysis was developed by Norquist and others (Norquist, 1982b, 1983, 1984, 1986; Halberstam et al., 1984), and was originally based on the OI procedures described in Bergman (1979) and McPherson et al. (1979). The normal mode initialization was based on the NMC NMI (Ballish, 1980) and is described in Norquist (1982a) and Tung (1983).

The design of the data assimilation experiment closely follows that described in Louis et al. (1987). An assimilation experiment consists of assimilation runs for two 7-day periods in the Special Observing Periods (SOPs) of the FGGE year: February 8 through 15, 1979, and June 17 through 24, 1979. Each assimilation run consists of a series of assimilation cycles using a 6 hour update cycle. Forecasts out to 4 days were produced from the initialized analyses at days 3, 5, and 7 of the assimilation runs. The assimilation period of the June assimilation run is identical with the independent sample for that month used in our colocation study. In February, the data assimilation period overlaps the dependent sample used in the regression study.

In this section we will present mostly differences between an assimilation experiment using the 3DNEPH based bogus RH data (referred to as NEPHSAT), and a control assimilation experiment using only the standard FGGE data set (STATSAT). The only moisture data used in STATSAT were radiosonde observations; for an in-depth discussion of the control run the reader is referred to Louis et al. (1987). The OI analysis program had undergone some minor changes between the STATSAT and NEPHSAT experiments: Changes were made to the quality-control procedures of drop-windsonde data, and to the procedures to solve the normal equations of the analysis program (Hoffman et al., 1988). This impact test differs from the one reported in Norquist (1988), in which bogus RH data was used to replace, rather than supplement, radiosonde measurements. The present OSE is designed to test whether the potential impact demonstrated by Norquist can be observed in a more realistic simulation of the operational environment.

The 3DNEPH based bogus RH data were generated for all half-mesh points located between 30°N and 50°N (since only those latitudes were used in the derivation of the regression equations); the error statistics for these data, which are needed as input to the moisture OI, were generated from the independent sample of the colocation study (see Appendix C for details).

The impact of the 3DNEPH data is most clearly seen in Fig. 7, which shows differences of analyzed RH at 850 hPa between NEPHSAT and STATSAT for the February assimilation run. The first analysis produced in February (February 8 at 06Z) shows differences to be essentially confined to the region influenced by the 3DNEPH data, i.e. 30°-50°N; an exception to this are the high latitude regions of both hemispheres, where sizable differences occur, due most probably to differences in the OI program between the two experiments. As is obvious from the succeeding panels in Fig. 7, the differences within the 30°-50°N latitude band as well as outside it grow with time. By 12Z on February 10, i.e. after 2 1/2 days of assimilation, the RH differences have spread over the entire globe, with the largest differences occurring at high latitudes; at that time, the region where bogus RH data were used in NEPHSAT is no longer visibly different from the rest of the world in these difference maps. This fairly rapid growth and spread of the RH differences indicate that the impact of the 3DNEPH data is within the noise level of the system, since the initial differences are clearly caused by both the different input data and the different analysis programs. During the early part of the assimilation run, however, the RH differences within the 30°-50°N latitude band can be related to the cloudiness data used in NEPHSAT. Plots of the 700 hPa cloud cover (Fig. 8), which is the predictor with most influence on bogus RH data for the 850 hPa level, reveal some areas of little cloud cover, particularly over the middle and East Atlantic, which correspond to areas where the 850 hPa NEPHSAT analyses are drier than the control.

The analyses of geopotential height are quite similar in the two experiments. During February, differences in the Northern Hemisphere are localized and of small amplitude (less than 100 m at 500 hPa, 150 m at 1000 hPa) throughout the entire assimilation run; in the Southern Hemisphere, large amplitude, small scale differences are evident near the pole, which are the result of the differences in the analysis programs and not related to the use of bogus RH data.

To assess the quality of the analyses and the forecasts produced from them, colocation statistics between the gridded fields and a set of verifying radiosonde observations were computed. The radiosonde observations used in the regression study were extracted from this data set. Fig. 9a shows the global rms error of relative humidity for NEPHSAT and STATSAT analyses and forecasts for February. Although the NEPHSAT errors are slightly smaller than those of STATSAT for most of the analyses, these differences (NEPHSAT-STATSAT) are smaller than the day-to-day variations of the analysis differences (analysis-RAOB). The forecasts do not indicate one to be superior to the other. The same general conclusions hold if one computes these statistics of just the Northern Hemisphere extratropics, or even just over North America (Fig. 9b and c), where the beneficial impact of the 3DNEPH data would be expected to be largest. Comparing the magnitudes of the rms errors shown in Fig. 9 with those of Table 6 might explain part of the reason for this apparent lack of improvement: the typical analysis errors are smaller than, and the typical 12-hour forecast errors are only slightly larger than, the colocation errors of the bogus RH data. The quality of the bogus RH data is thus comparable to that of the first guess, resulting in only a small positive impact even in radiosonde-void regions. The generally inconclusive results of the 850 hPa radiosonde statistics hold for other levels, as well.

The results from the June assimilation run are quite similar to those for February. Difference maps of RH, shown in Fig. 10, reveal the impact of the cloud data in the first analysis time periods, along with large differences near the South Pole, which are related to the different analysis program. As was the case in February, the differences spread quickly over the entire globe, until at day 2 1/2 the region with cloud data input is no longer distinguishable from the rest of the globe. The 3DNEPH cloud cover data shown in Fig. 11 shows some features that can be related to the RH differences, in particular an area of positive RH differences and high values of cloud cover at 150°W.

The radiosonde statistics for the June assimilation also give no clear indication that the NEPHSAT forecasts have smaller RH errors than those of STATSAT.

6. Summary and Conclusions

Based on a colocation study of radiosonde RH measurements and 3DNEPH cloudiness data over North America, we developed regression equations for the first two EOFs of relative humidity. The regression equations predict the coefficients of EOF-All, i.e. the EOFs based on the mean RH profile computed for all latitudes between 20° and 50°N, from the vertically compacted, horizontally averaged 3DNEPH data set. The regression equations performed better than existing level-to-level methods for inferring RH from cloud cover. Although the verification was performed over the same geographical region for which the regression equations were developed, and no attempt was made to tune the existing methods to improve their performance, this comparison establishes the feasibility of the approach.

The utility of the bogus RH data for operational data assimilation was investigated in a global observing system experiment, in which bogus RH data were supplied to the moisture analysis in the 30°-50°N latitude belt. The impact of the 3DNEPH data was evident in analyses in the early part of the assimilation runs; at later times, it was more difficult to separate the effects of the bogus RH data from other differences between the NEPHSAT and control OSE. Comparisons of the analyzed and forecast RH with verifying radiosonde data did not indicate a measurable positive impact of the bogus RH data. The inconclusive results from this OSE should not be regarded as definitive, however; we address the reasons for the lack of positive impact, and suggest possible extensions to the present study in the following.

One of the obvious shortcomings of the present OSE is the limited geographical extent of the bogus RH data. In future studies, the RH profile approach could be extended to produce a global bogus RH data set by repeating the regression study performed here for different regions of the globe. Different EOFs, and different regression equations would then be used in different regions.

Other limitations of the OSE are related to the data assimilation system itself. Among those the relatively coarse resolution of the analysis and forecast, the use of an adiabatic NMI, and of a very simple moist physics package in the GSM are the most significant obstacles to an effective assimilation of moisture data. There are several potential remedies to these short-

comings: using a diabatic NMI, in conjunction with a moisture spinup procedure as suggested in Donner (1988) would minimize the rejection of initial moisture data by the forecast model. Improvements to the physics package of the GSM are also necessary to limit the error growth during the assimilation cycle and the longer range forecasts produced from the analyses; the physics package currently being implemented and tested by AFGL is expected to improve this aspect of the GDAS.

Perhaps the most serious limitation to the usefulness of the data are the relatively large observation errors of the bogus RH, which are larger than the globally averaged errors of the current RH analyses. Even with the current error levels, however, some beneficial impact should be realizable in otherwise data-void areas. It may also be possible to reduce the errors of the bogus RH data with changes in the regression approach, such as the definition of the EOFs or the preprocessing of the cloud data. However, for significant reductions of the observation errors, it will be necessary to take account of the fact that there is no one-to-one correspondence between relative humidity and cloud cover, and to include other atmospheric parameters (e.g., static stability, vertical motion) in the problem.

7. Acknowledgements

The authors are indebted to Don Norquist for providing the quality-checked RAOB data set. We benefited from helpful discussions with Don Norquist and Ken Mitchell of AFGL, and Jean-Francois Louis and Ron Isaacs of AER. Thanks are also due to M. Mickelson for running the impact study, D. Coombs for producing the plots, and L. Struz for preparing the manuscript.

8. References

- Ballish, B., 1980: Initialization, theory and application to the NMC spectral model. Ph.D. Thesis, Dept. of Meteorology, University of Maryland, College Park, MD, 151 pp.

- Bergman, K.H., 1979: Multivariate analysis of temperature and winds using optimum interpolation. Mon. Wea. Rev., 107, 1423-1444.
- Brenner, S., C.-H. Yang, and S.Y.K. Yee, 1982: The AFGL spectral model of the moist global atmosphere: Documentation of the baseline version. AFGL-TR-82-0393, AFGL, Hanscom AFB, MA, 65 pp. [NTIS ADA129283].
- Brenner, S., C.-H. Yang, and K. Mitchell, 1984: The AFGL global spectral model: Expanded resolution baseline version. AFGL-TR-84-0307, AFGL, Hanscom AFB, MA, 72 pp. [NTIS ADA160370].
- Chu, R., and D. Parrish, 1977: Humidity analysis for operational prediction models at the National Meteorological Center. NOAA NMC Office Note 140, 32 pp. [Available from NMC, 5200 Auth Road, Washington, DC 20233].
- Donner, L.J., 1988: An initialization for cumulus convection in numerical weather prediction models. Mon. Wea. Rev., 116, 377-385.
- Draper, N.R. and H. Smith, 1966: Applied Regression Analysis. Wiley, New York, 407 pp.
- Efroymson, M.A., 1960: Multiple regression analysis. In Mathematical Methods for Digital Computers, A. Ralston and H.S. Wilf, editors, John Wiley & Sons, New York, 191-203.
- Fye, F.K., 1978: The AFGWC Automated Cloud Analysis Model. AFGWC Tech. Memo. 78-002, 97 pp. [Available from AFGWC/SDDN, Offutt AFB, NE 68113-5001].
- Geleyn, J.-F., 1981: Some diagnostics of the cloud/radiation interaction in ECMWF forecasting model. In ECMWF Workshop on Radiation and Cloud/Radiation Interaction in Numerical Modeling, Reading, UK.
- Halberstam, I., C. Johnson, D.C. Norquist, and S.-L. Tung, 1984: Two methods of global data assimilation. AFGL-TR-84-0260, AFGL, Hanscom AFB, MA, 91 pp. [NTIS ADA155981].

- Hoffman, R.H., M. Mickelson and T. Nehrkorn, 1988: Enhancements to the AFGL statistical analysis program (ASAP) for the global multivariate analysis of heights and winds. AFGL-TR-88-0279, AFGL, Hanscom AFB, MA.
- Kanamitsu, M., 1984: Data processing, selection, and quality control as used at the Japanese Meteorological Agency. In ECMWF Workshop on the Use and Quality Control of Meteorological Observations, Reading, UK.
- Louis, J.-F., R.H. Hoffman, T. Nehrkorn, T. Baldwin and M. Mickelson, 1987: Baseline observing system experiments using the AFGL global data assimilation system. AFGL-TR-87-0327, AFGL, Hanscom AFB, MA. [NTIS ADA188861]
- McPherson, R.D., K.H. Bergman, R.E. Kistler, G.E. Rasch, and D.S. Gordon, 1979: The NMC operational global data assimilation system. Mon. Wea. Rev., 107, 1445-1461.
- Neter, J. and W. Wasserman, 1975: Applied Linear Statistical Models. Richard D. Irwin, Inc., Homewood, Illinois, 842 pp.
- Norquist, D.C., 1982a: Adaptation and application of NMC nonlinear normal mode initialization scheme. In Objective Analysis and Prediction Techniques, A.M. Gerlach (Ed.), AFGL-TR-82-0394, AFGL, Hanscom AFB, MA, pp. 21-32 [NTIS ADA131465].
- Norquist, D.C., 1982b: Development of objective analysis schemes using the method of optimum interpolation. In Objective Analysis and Prediction Techniques, A.M. Gerlach (Ed.), AFGL-TR-82-0394, AFGL, Hanscom AFB, MA, pp. 51-55 [NTIS ADA131465].
- Norquist, D.C., 1983: Development and testing of a multivariate global statistical analysis system. In Objective Analysis and Prediction Techniques, A.M. Gerlach (Ed.), AFGL-TR-83-C333, AFGL, Hanscom AFB, MA, pp. 10-48 [NTIS ADA142441].

- Norquist, D.C., 1984: Users guide for optimum interpolation method of global data assimilation. AFGL-TR-84-0290, AFGL, Hanscom AFB, MA, 67 pp.. [NTIS ADA155929].
- Norquist, D.C., 1986: Alternative forms of moisture information in 4-D data assimilation. AFGL-TR-86-0194, AFGL, Hanscom AFB, MA, 71 pp.. [NTIS ADA179792].
- Norquist, D.C., 1988: Alternative forms of humidity information in global data assimilation. Mon. Wea. Rev., 116, 452-471.
- Rasmussen, R.G., 1982: Some techniques for the objective analysis of humidity for regional scale numerical weather prediction. NCAR Cooperative Thesis No. 67, Boulder, CO.
- Saito, K., and A. Baba, 1988: A statistical relation between relative humidity and the GMS observed cloud amount. J. Met. Soc. Japan, 66, 187-192.
- Sela, J.G., 1980: Spectral modeling at the National Meteorological Center. Mon. Wea. Rev., 108, 1279-1292.
- Tibaldi, S., 1982: The ECMWF humidity analysis and its general impact on global forecasts and on the forecast in the Mediterranean area in particular. Rivista di Meteorologia Aeronautica, 42, 309-328 [portion used in this report also available in Atmospheric Water Vapor, 1980, Academic Press, ISBN 0-12-208440-3].
- Tung, S.L., 1983: Generalization and application of Machenhauer initialization scheme. In Objective Analysis and Prediction Techniques, A.M. Gerlach (Ed.), AFGL-TR-83-0333, AFGL, Hanscom AFB, MA [NTIS ADA142441].

Appendix A: Existing cloudiness to humidity conversion techniques

Three existing methods of estimating relative humidity on mandatory pressure levels were implemented in our colocation study. All three methods were used with the horizontal mean values of the vertically compacted 3DNEPH cloud cover corresponding to each mandatory pressure level. In cases of zero cloud cover, the smaller of the critical relative humidity and climatology was used as the bogus RH. In the case of the Tibaldi and inverse ECMWF methods, which require knowledge of the surface pressure, the surface pressure of the co-located radiosonde observation was used.

A.1 Tibaldi Method

Equation (1) was applied to four humidity layers $j=1,4$, where the layers are defined by $p_j \leq p \leq p_{j+1}$, with the p_j and the coefficients M_j , A_j given in Table A.1.

Table A.1: Parameters of the Tibaldi method

j	p_j	M_j	A_j
1	p_{surface}	.80	.20
2	$p_1 - 50 \text{ hPa}$.75	.15
3	$p_2 - (p_2 - p_5)/3$.60	.15
4	$p_3 - (p_2 - p_5)/3$.55	.10
5	300 hPa		

A.2 AFGWC method

This method is based on empirical relationships between the cloud cover and the condensation pressure spread (CPS), which are defined for the mandatory pressure levels 850, 700, 500, and 300 hPa (Table A.2; see Norquist, 1988, for a graphical display of the curves). We used the 850 hPa curve at 1000 hPa and the average of the 500 and 300 hPa curves at 400 hPa. The CPS is related to the dew-point depression (DPD) through the approximate relationship:

$$\text{DPD} = \text{CPS} / (a_0 + a_1(p/p_0) + a_2(p/p_0)^2) ,$$

where $p_0=1000$ hPa, $a_0=4.9$ hPa/K, $a_1=.93$ hPa/K, and $a_2=-9$ hPa/K. To convert the dew point depression to relative humidity, the temperatures of the colocated radiosonde were used.

Table A.2

Values of condensation pressure spread (CPS, in hPa) as a function of cloud cover, for the mandatory pressure levels at 850, 700, 500, and 300 hPa. The entry in a particular row and column of these tables is the CPS corresponding to the cloud cover (in %) obtained by adding the labels of the row and column; e.g., the entry in the first column of the second row corresponds to 11% cloud cover.

P = 850 hPa

	1	2	3	4	5	6	7	8	9	10
0	114.2	111.3	108.5	105.8	103.2	100.7	98.2	95.7	93.2	90.7
10	88.0	85.2	82.5	80.0	77.5	75.2	73.2	71.3	69.6	68.1
20	66.6	65.2	63.9	62.5	61.1	59.6	58.0	56.6	55.3	54.2
30	53.1	52.0	51.0	50.0	49.1	48.3	47.4	46.6	45.8	45.0
40	44.2	43.4	42.6	41.8	41.0	40.2	39.4	38.6	37.8	37.0
50	36.2	35.4	34.8	34.4	34.0	33.7	33.3	32.9	32.6	32.3
60	31.9	31.5	31.1	30.7	30.2	29.7	29.2	28.7	28.3	27.9
70	27.5	27.1	26.7	26.2	25.7	25.2	24.7	24.2	23.7	23.1
80	22.5	21.9	21.2	20.4	19.5	18.5	17.5	16.6	15.6	14.6
90	13.6	12.5	11.4	10.3	9.1	7.6	5.9	4.3	2.7	1.0

P = 700 hPa

	1	2	3	4	5	6	7	8	9	10
0	107.2	105.0	102.8	100.8	98.9	97.5	96.2	95.0	93.9	92.8
10	91.7	90.6	89.5	88.3	86.8	85.0	83.0	81.0	79.0	77.0
20	75.2	73.4	71.6	69.9	68.3	66.8	65.3	63.9	62.5	61.2
30	60.0	58.7	57.4	56.2	55.1	54.0	52.9	51.8	50.7	49.7
40	48.8	47.9	47.0	46.1	45.3	44.5	43.7	42.8	42.0	41.3
50	40.7	40.3	39.9	39.6	39.3	38.9	38.6	38.2	37.9	37.6
60	37.2	36.8	36.4	36.0	35.6	35.1	34.7	34.2	33.8	33.3
70	32.8	32.2	31.7	31.2	30.7	30.2	29.7	29.1	28.5	28.0
80	27.5	27.0	26.4	25.8	25.1	24.3	23.4	22.4	21.3	20.0
90	18.7	17.4	16.0	14.6	13.2	11.6	9.8	7.4	4.5	1.0

P = 500 hPa

	1	2	3	4	5	6	7	8	9	10
0	100.2	99.4	98.6	97.8	97.0	96.2	95.4	94.5	93.7	92.9
10	92.1	91.4	90.6	89.5	88.3	87.0	85.7	84.3	82.9	81.2
20	79.4	77.7	76.0	74.4	72.8	71.5	70.5	69.5	68.5	67.6
30	66.7	65.8	65.0	64.1	63.1	62.1	61.1	59.8	58.2	56.4
40	54.4	52.6	51.3	50.3	49.6	48.9	48.3	47.6	46.9	46.2
50	45.5	44.8	44.1	43.4	42.7	42.0	41.3	40.8	40.4	40.0
60	39.6	39.2	38.8	38.4	38.0	37.6	37.2	36.8	36.3	35.9
70	35.5	35.0	34.6	34.2	33.7	33.3	32.8	32.1	31.4	30.7
80	29.9	29.0	28.1	27.2	26.2	25.1	24.0	22.9	21.8	20.7
90	19.6	18.3	16.9	15.3	13.4	11.4	9.4	7.4	5.0	1.0

P = 300 hPa

	1	2	3	4	5	6	7	8	9	10
0	96.2	95.4	94.6	93.9	93.1	92.3	91.4	90.5	89.7	88.9
10	88.0	87.1	86.4	85.7	85.1	84.5	83.7	82.8	81.7	80.4
20	79.0	77.5	75.9	74.3	72.9	71.6	70.4	69.4	68.4	67.5
30	66.7	65.8	64.9	64.0	63.1	62.2	61.2	59.8	58.2	56.4
40	54.8	53.5	52.2	51.1	50.1	49.2	48.4	47.6	46.9	46.2
50	45.5	44.8	44.1	43.4	42.7	42.0	41.3	40.8	40.4	40.0
60	39.6	39.2	38.8	38.4	38.0	37.6	37.2	36.8	36.3	35.8
70	35.4	35.0	34.6	34.2	33.8	33.3	32.8	32.2	31.5	30.6
80	29.7	28.8	27.9	27.0	26.1	25.1	24.0	23.0	21.9	20.7
90	19.5	18.3	16.9	15.3	13.4	11.4	9.4	7.4	5.0	1.0

A.3 Inverse ECMWF method

Equation (2) was inverted to yield

$$RH = RH_c + \eta^{1/2} / (1 - RH_c)$$

where RH (relative humidity) and η (cloud cover) are dimensionless, and RH_c , the critical RH is given by

$$RH_c = 1 - \alpha\sigma' (1 - \sigma') (1 + \beta(\sigma' - 1/2)) ,$$

with $\sigma' = p/p_s$, $\alpha=2$, and $\beta=3^{1/2}$.

Appendix B: The 3DNEPH data base

The 3DNEPH analysis as described in Fye (1978) provides the following data at each grid point: terrain height (in tens of meters), total cloud cover (to the nearest percent), cloud cover for 15 layers (to the nearest 5 percent), minimum cloud base and maximum cloud top (in WMO code 1677), cloud type for low, middle, and high clouds (in AFGWC codes), and present weather (in WMO code 4677, divided by 10). The vertical structure of the 3DNEPH layers is shown in Table B.1.

Table B.1
Vertical Structure of the 3DNEPH layers.
Adapted from Fye (1978)

Layer Number	Upper Boundary Height (m)		Compacted Layer
	Surface		
1	46	AGL	LOW
2	91	AGL	LOW
3	183	AGL	LOW
4	305	AGL	LOW
5	610	AGL	1000 hPa
6	1067	AGL/MSL	1000 hPa
7	1524	MSL	850 hPa
8	1981	MSL	850 hPa
9	3048	MSL	700 hPa
10	4267	MSL	700 hPa
11	5486	MSL	500 hPa
12	6706	MSL	500 hPa
13	7925	MSL	400 hPa
14	10668	MSL	300 hPa
15	16764	MSL	300 hPa

Of these, all but the present weather information were used in the collocation study. Some recoding was performed to allow a quantitative analysis. Cloud cover values that denoted "thin clouds" were set to zero unless they were at layers 13 or above. The minimum cloud base and maximum cloud top codes were converted to heights in meters. The cloud type information was recoded such that missing information, i.e. no clouds, was assigned a value of zero, cumuliform cloud types a value of 10, mixed cumuliform and stratiform cloud types a value of 15, and stratiform cloud types a value of 20. Vertically compacted cloud cover values were computed assuming maximum overlap; the 3DNEPH layers assigned to the compacted cloud cover layers are shown in Table B.1. Except for the low cloud value, the vertically compacted cloud cover values were computed by assigning the terrain-following (AGL) 3DNEPH layer cloud cover values to the corresponding constant altitude (MSL) layers in the case of elevated terrain. In order to compute horizontal averages, the terrain-following 3DNEPH layer values were also remapped to equivalent constant altitude layers. Horizontal averages and standard deviations were computed over 5x5 grid points centered around the grid point closest to each colocated radiosonde station, using a 1-2-1 weighted average in both directions. For the collocation study data was extracted from 3DNEPH boxes 43, 44, 45. For central gridpoints near box boundaries, averages and standard deviations were computed over fewer than 25 points. The OSE used data in all 3DNEPH boxes between 30°N and 50°N.

Appendix C: Error statistics of the bogus RH data

The bogus RH data was used in the moisture OI, with error statistics that were estimated from the independent sample of the colocation study. Different error statistics were used for February and June. The observational error standard deviations (σ_N) were estimated from the RMS colocation errors (σ_c , viz. Tables 6 and 8), using the relationship (Bergman, 1978, eq 12):

$$\sigma_N^2 = \sigma_c^2 - \sigma_R^2$$

with an assumed RAOB error (σ_R) of 5%. This same RAOB error was used in the moisture OI.

The vertical error correlations (Tables C.1 and C.2) were computed by first computing a colocation error covariance matrix, using the colocation errors of the EOF coefficients and the RH EOFs; the assumed RAOB error covariance matrix was then subtracted, and correlations computed from the resulting matrix. RAOB errors were assumed to be uncorrelated in the vertical, resulting in an error matrix with nonzero elements only along the diagonal, with a magnitude corresponding to a 5% error. The colocation error covariance between two levels k, ℓ is related to the EOF coefficient errors by the following relationship:

$$\overline{(\epsilon'_{r,k} \epsilon'_{r,\ell})} = \sum_{m=1}^{15} \overline{(\epsilon'_{c,m})^2} E_{m,k} E_{m,\ell}$$

where $\epsilon_{r,k}$ denotes the colocation error of RH at level k , $\epsilon_{c,m}$ the colocation error of EOF coefficient m , and $E_{m,k}$ the amplitude of EOF m at level k . It was assumed here that the coefficient errors of two EOFs are uncorrelated (which is strictly true only for the dependent sample). The above equation follows from the property of the EOFs that any RH profile may be written as:

$$RH_k = \overline{RH}_k + \sum_{m=1}^{15} \epsilon_m E_{m,k}$$

TABLE C.1

The vertical error correlations of the bogus RH data for the 15 sigma levels for February.

LEVEL	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.000														
2	.728	1.000													
3	.516	.848	1.000												
4	.396	.606	.840	1.000											
5	.276	.402	.593	.844	1.000										
6	.182	.276	.401	.599	.841	1.000									
7	.115	.203	.270	.406	.601	.841	1.000								
8	.076	.154	.171	.271	.426	.630	.875	1.000							
9	.051	.105	.111	.177	.301	.464	.678	.888	1.000						
10	.023	.075	.074	.116	.205	.349	.519	.702	.888	1.000					
11	.002	.028	.030	.070	.133	.243	.375	.521	.675	.855	1.000				
12	-.012	-.001	-.019	.021	.080	.181	.295	.409	.531	.680	.885	1.000			
13	-.037	-.035	-.053	-.022	.025	.118	.220	.317	.409	.524	.702	.863	1.000		
14	-.066	-.056	-.068	-.044	-.002	.082	.165	.240	.318	.412	.555	.689	.889	1.000	
15	-.076	-.084	-.090	-.061	-.028	.045	.125	.191	.257	.324	.466	.580	.755	.916	1.000

Table C.2

The vertical correlations of the bogus RH data for the 15 sigma levels for June.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.000														
2	.774	1.000													
3	.536	.883	1.000												
4	.315	.574	.826	1.000											
5	.180	.302	.459	.731	1.000										
6	.103	.163	.257	.416	.759	1.000									
7	.071	.121	.167	.246	.460	.773	1.000								
8	.072	.088	.122	.158	.293	.505	.798	1.000							
9	.081	.085	.085	.085	.169	.335	.559	.807	1.000						
10	.071	.088	.080	.068	.121	.234	.392	.566	.791	1.000					
11	.083	.111	.112	.104	.111	.168	.255	.377	.515	.789	1.000				
12	.092	.124	.127	.126	.106	.119	.153	.249	.348	.546	.786	1.000			
13	.104	.111	.104	.133	.127	.109	.133	.182	.223	.337	.500	.747	1.000		
14	.068	.056	.067	.106	.112	.109	.125	.141	.157	.214	.316	.513	.746	1.000	
15	.075	.053	.068	.099	.100	.112	.109	.116	.137	.177	.221	.360	.518	.779	1.000

where \overline{RH}_k is the climatological mean profile, and e_m the coefficient of the m^{th} EOF.

Finally, horizontal error correlation functions were derived from collocation errors. Error correlations were computed for distance bins of 200 km width. Results showed anomalously high values for the surface level out to large distances which were caused by a diurnally varying bias. Because of this bias, bogus RH values at the surface were not used, and curves were fitted to error correlation values that were averaged over the remaining levels. The fitted curves were of the form $\exp(-d/k)$, where d is the separation distance, and k is 290 km (230 km) for February (June).

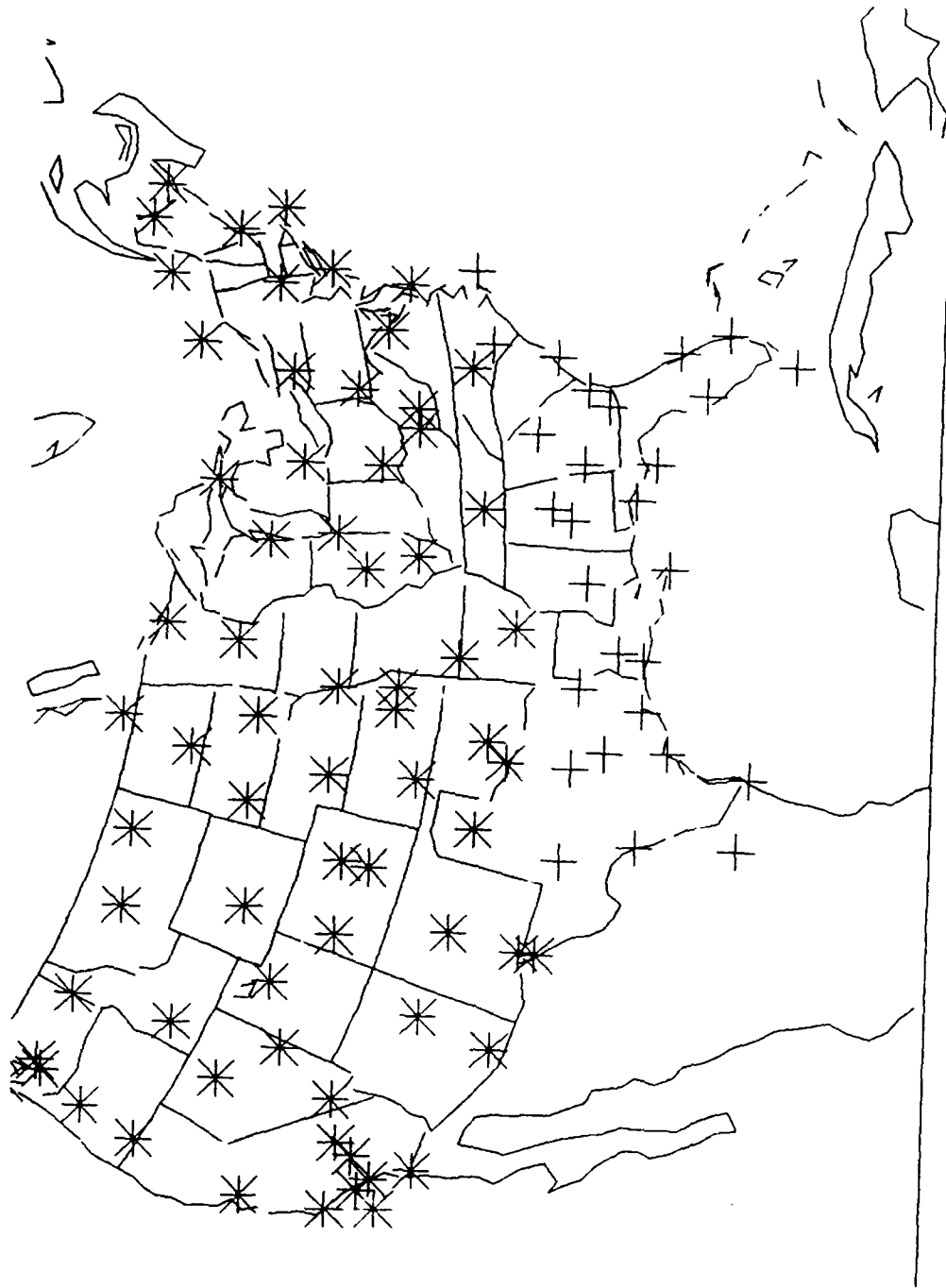


Fig. 1: Map of radiosonde stations used in the regression study. Crosses denote stations at which only relative humidity data were used, stars denote stations with relative humidity and 3DNEPH cloud data.

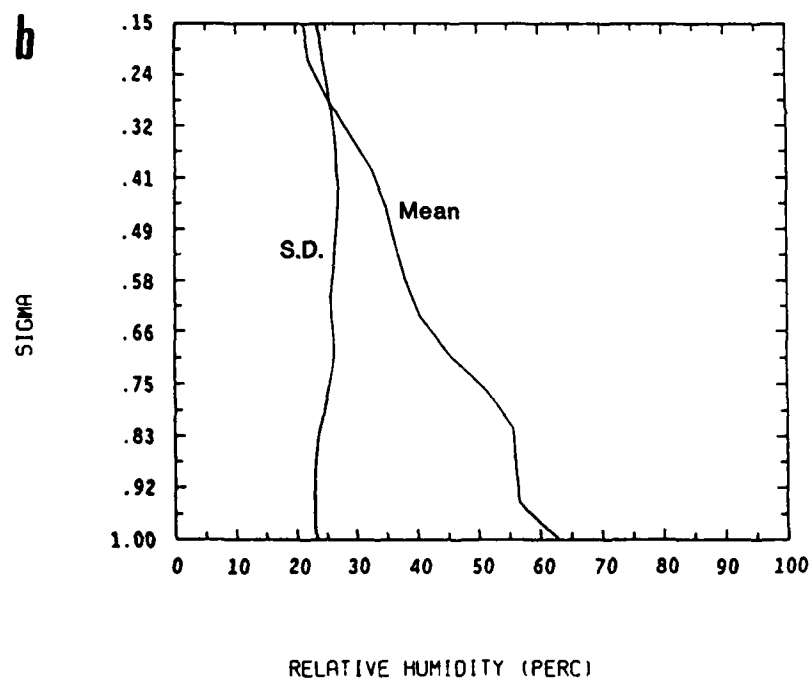
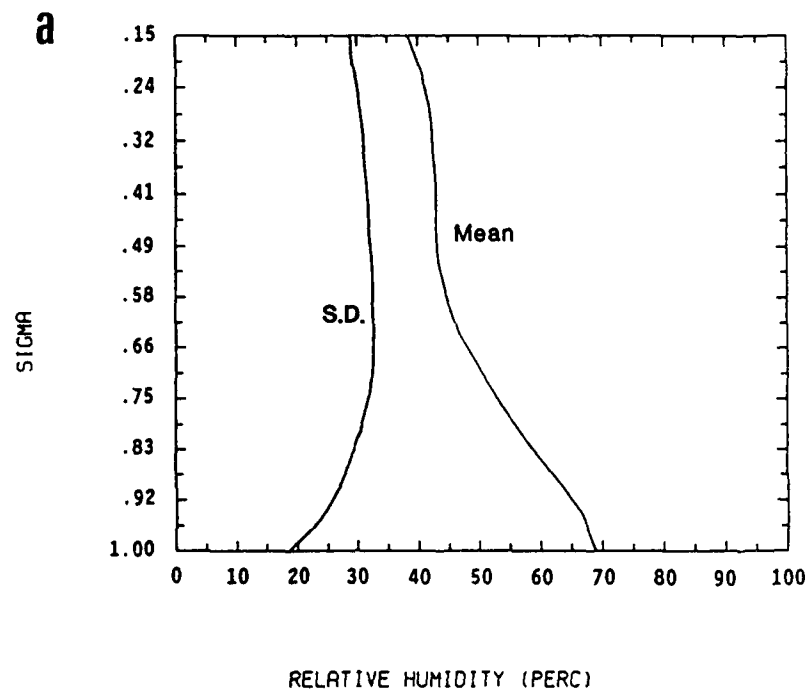


Fig. 2: Mean and standard deviation relative humidity profiles for EOF-All for February (a) and June (b).

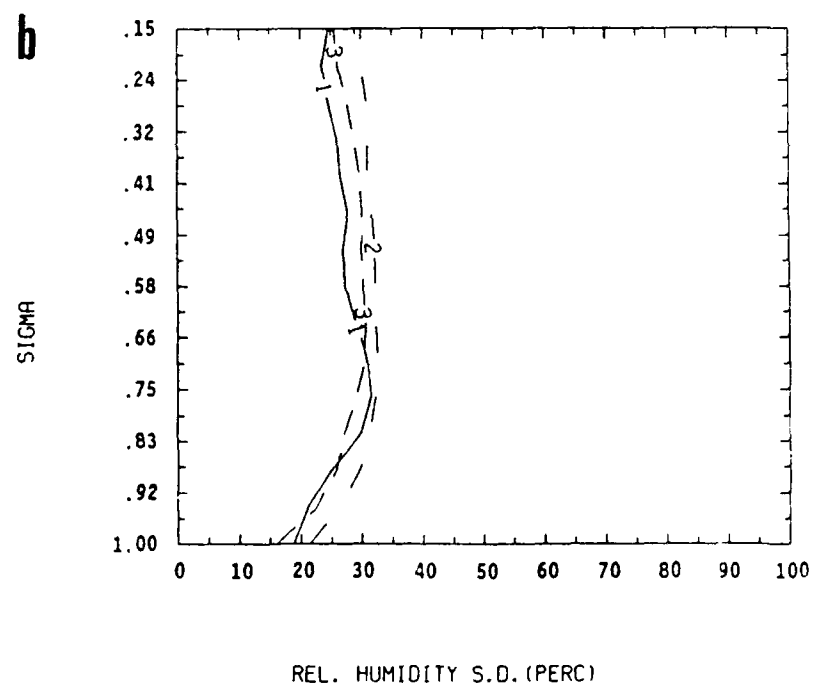
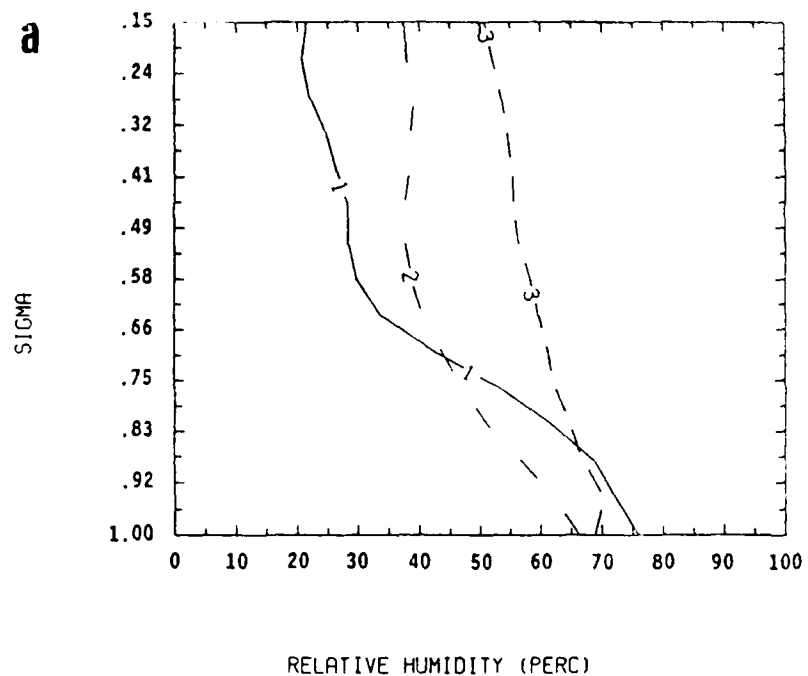


Fig. 3: Mean (a,c) and standard deviation (b,d) relative humidity profiles for EOF-Bands for February (a,b) and June (c,d). The curve labels (1,2,3) refer to the three latitude bands (20°N-30°N, 30°N-40°N, 40°N-50°N).

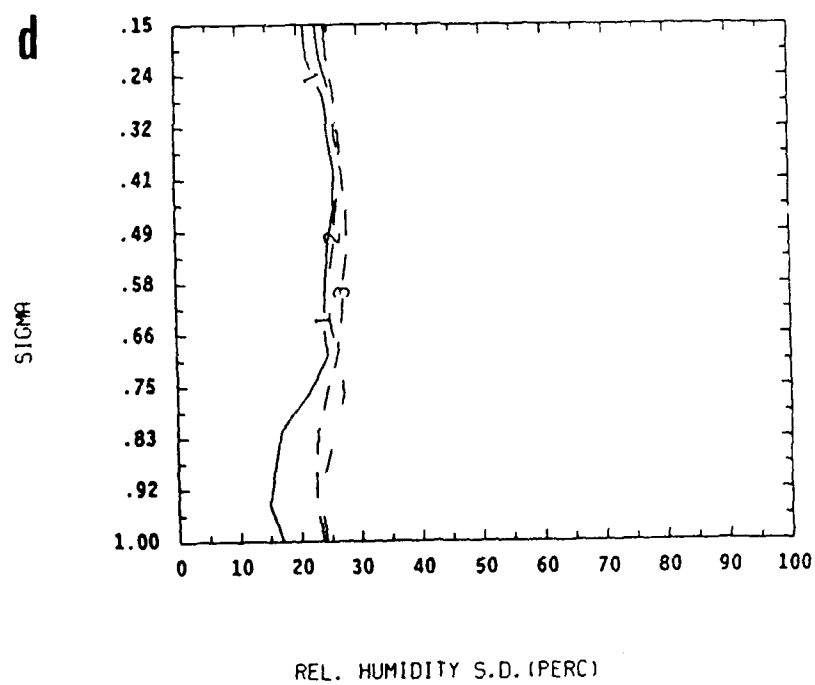
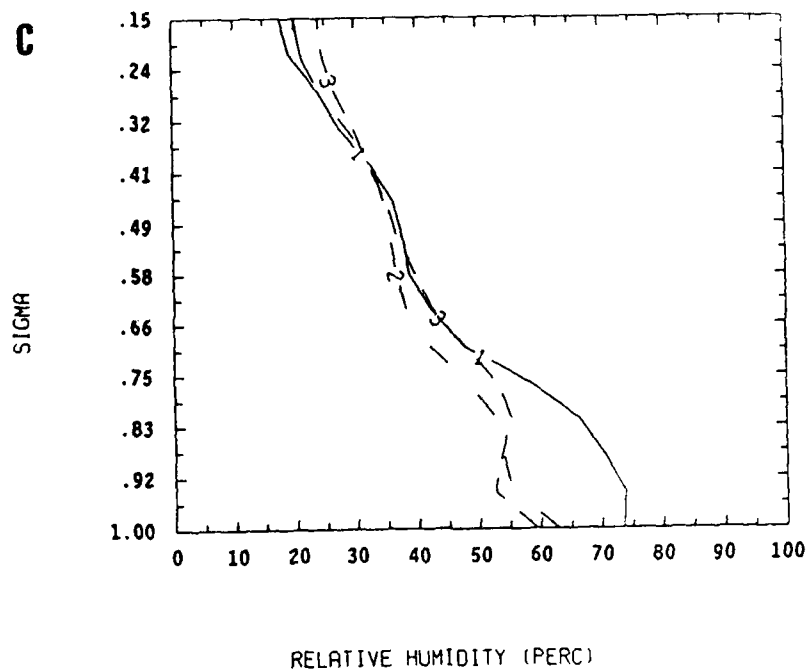


Fig. 3: Continued.

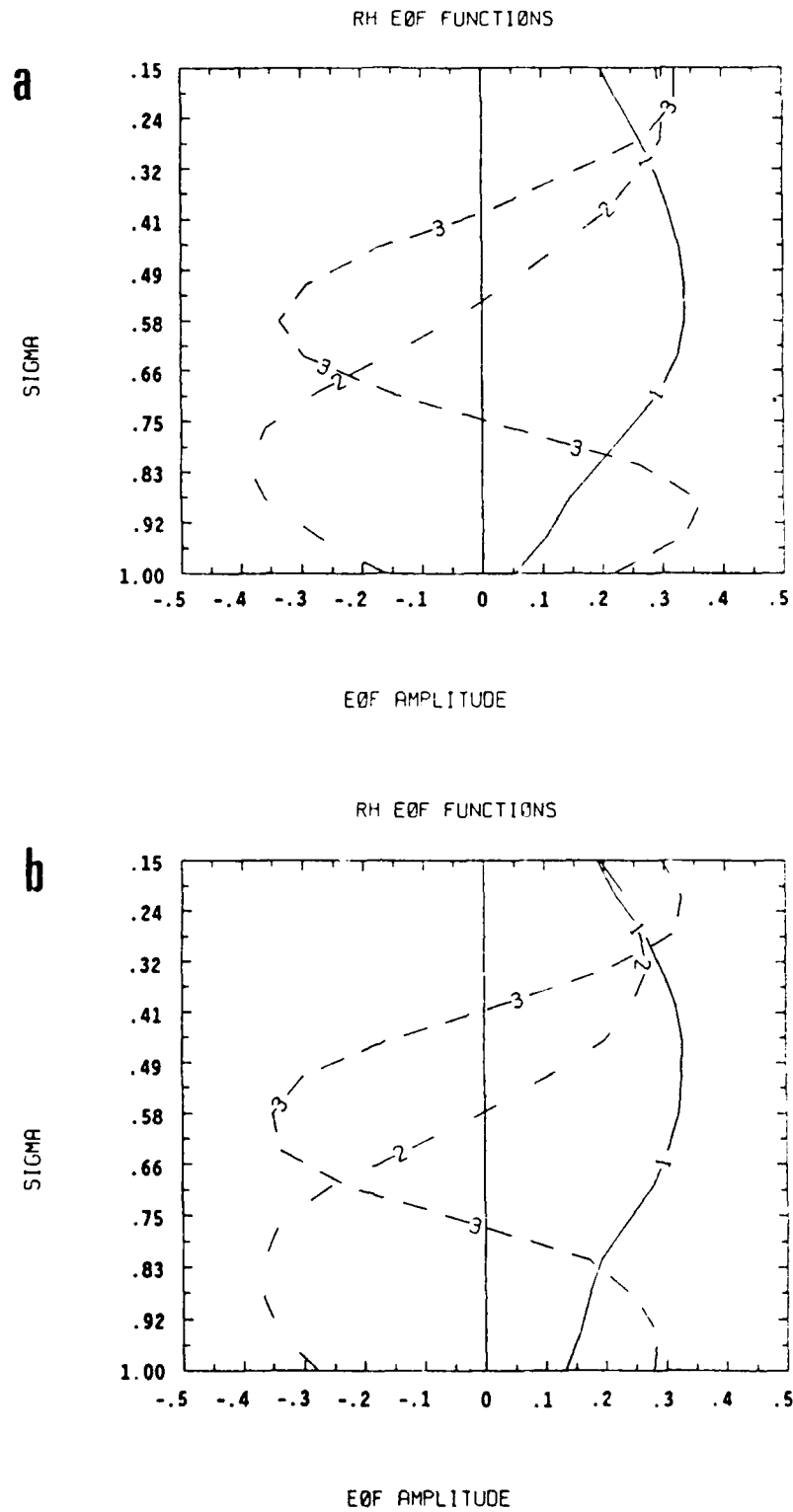


Fig. 4: Profiles of the first three EOFs for EOF-All for February (a) and June (b).

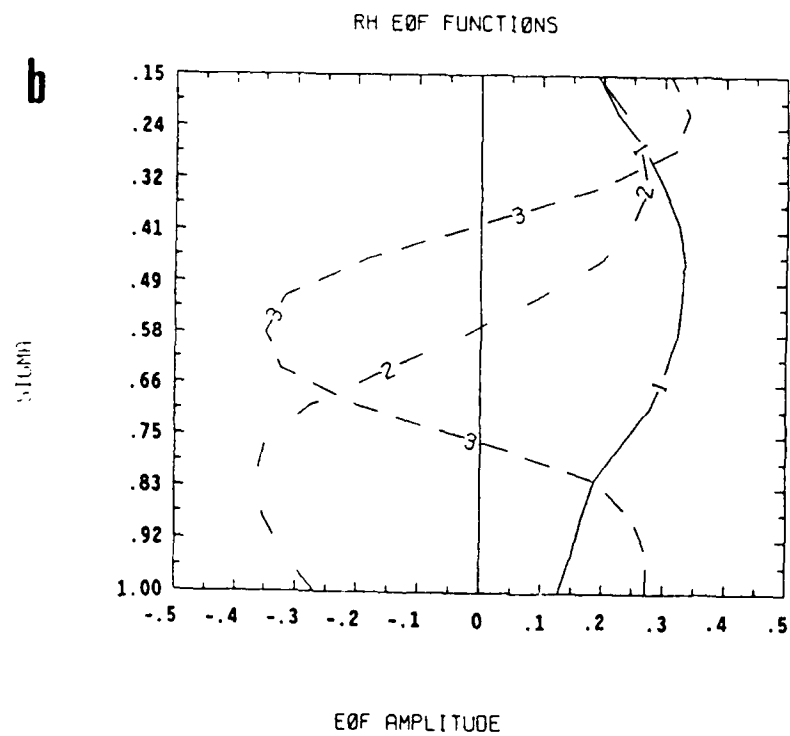
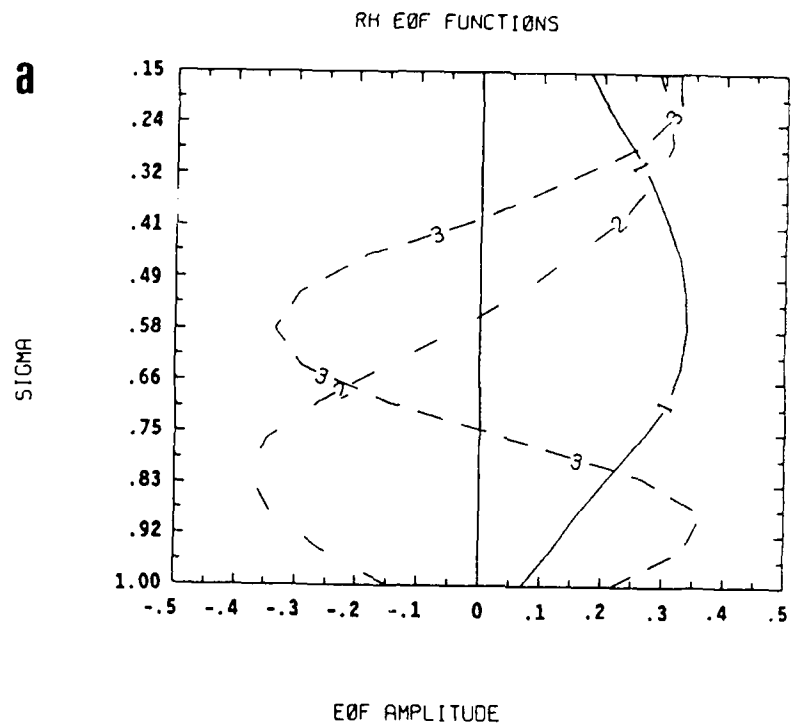


Fig. 5: As Fig. 4, but for EOF-Bands.

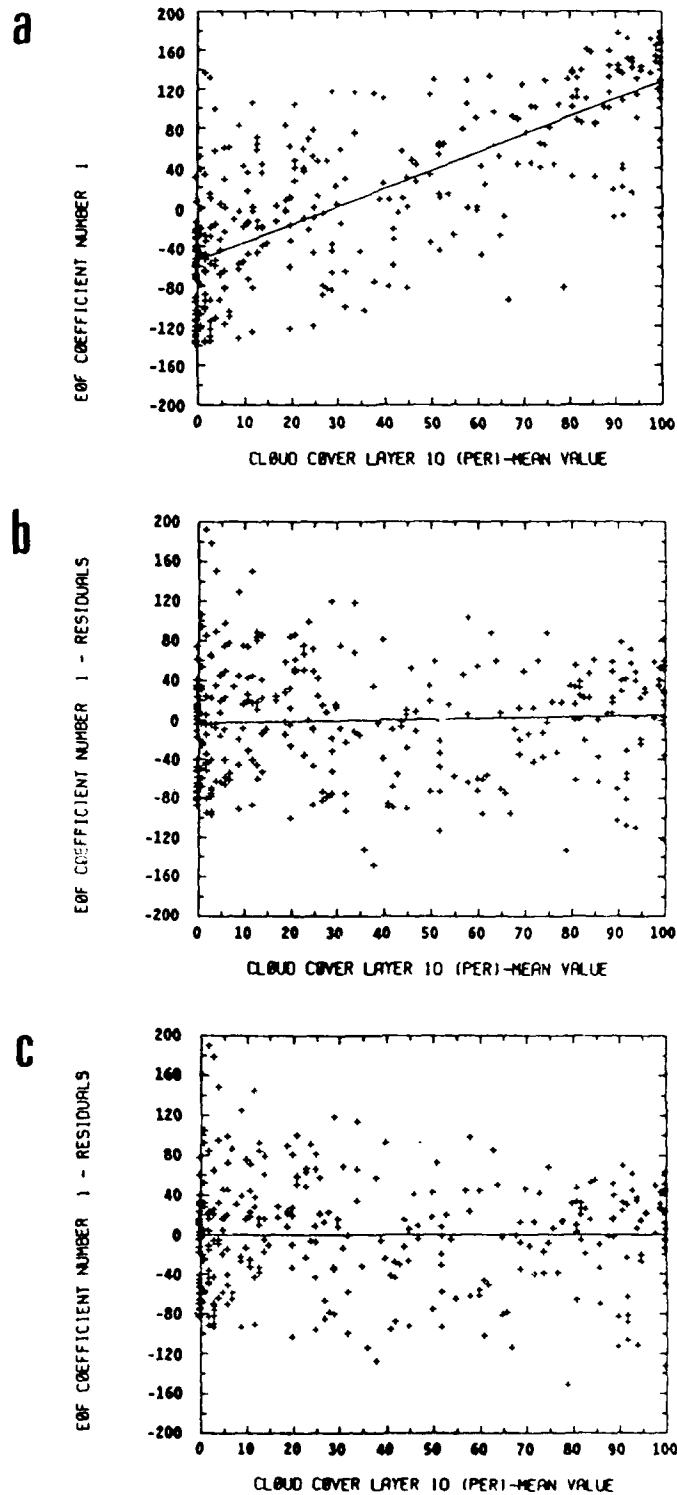
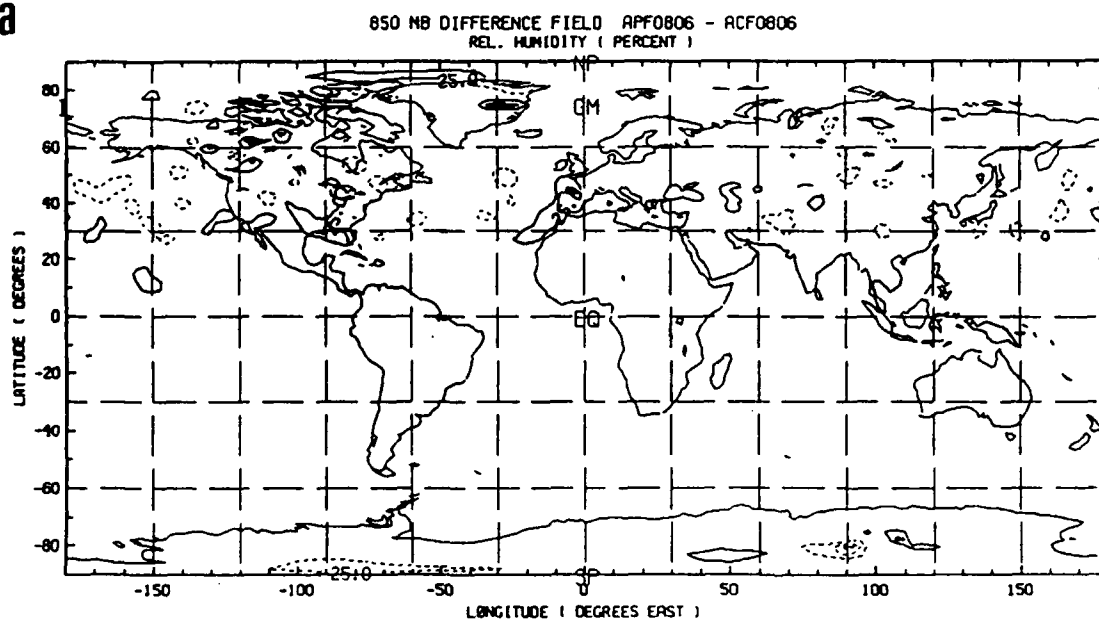


Fig. 6: Scatterplots of EOF coefficient 1 values (a) and residuals for regression equation using either compacted 3DNEPH data set (b) or full 3DNEPH data set (c) versus layer 10 cloud cover values for February.

a



b

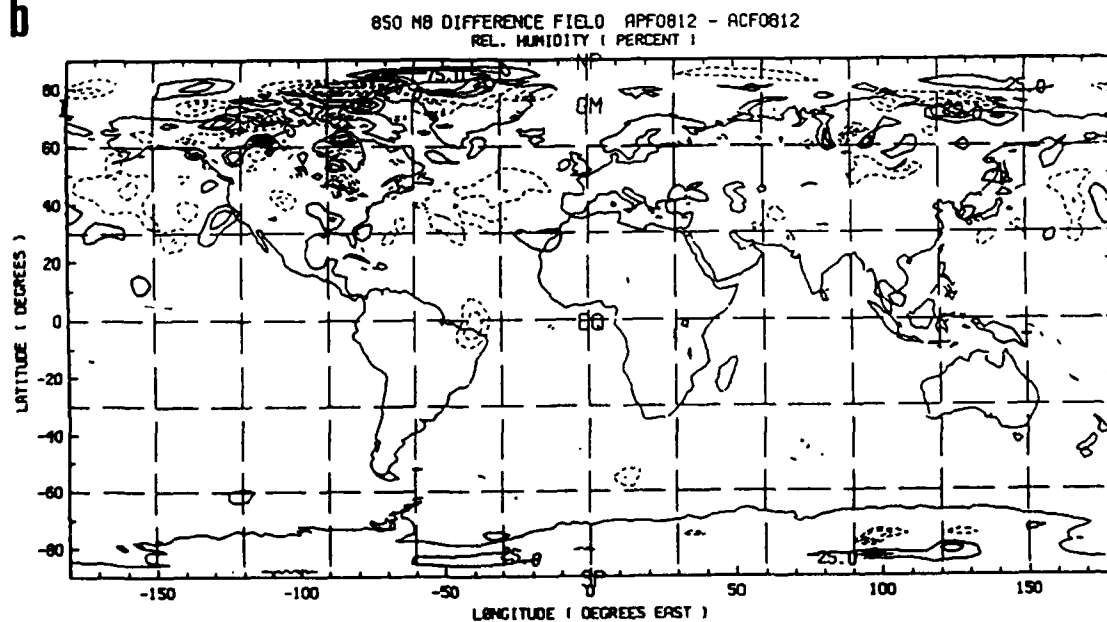
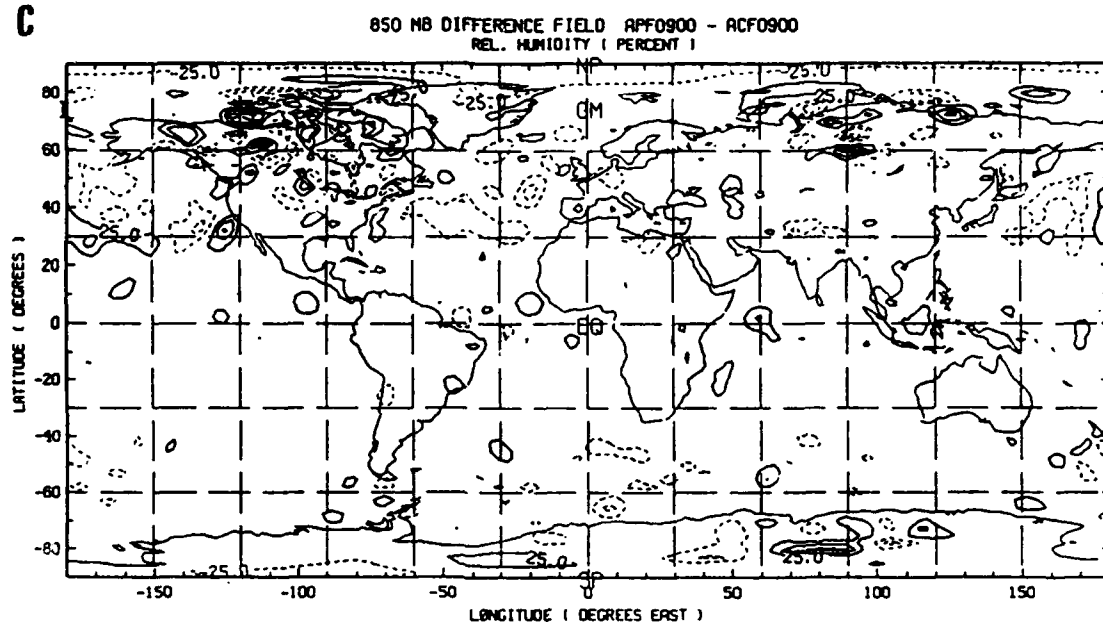


Fig. 7: Contour plots of RH differences between NEPHSAT and STATSAT analyses for February 8, 06Z (a), February 8, 12Z (b), February 9, 00Z (c), and February 10, 12Z (d). Contours are drawn every 25%, the zero contour is suppressed, and negative values are dashed.

C



d

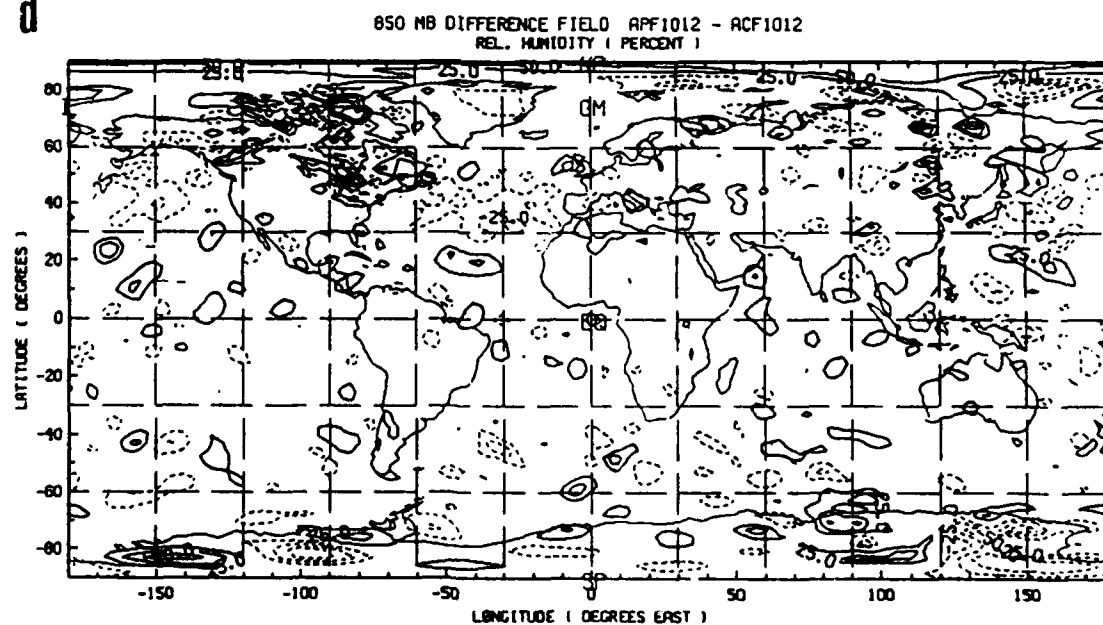
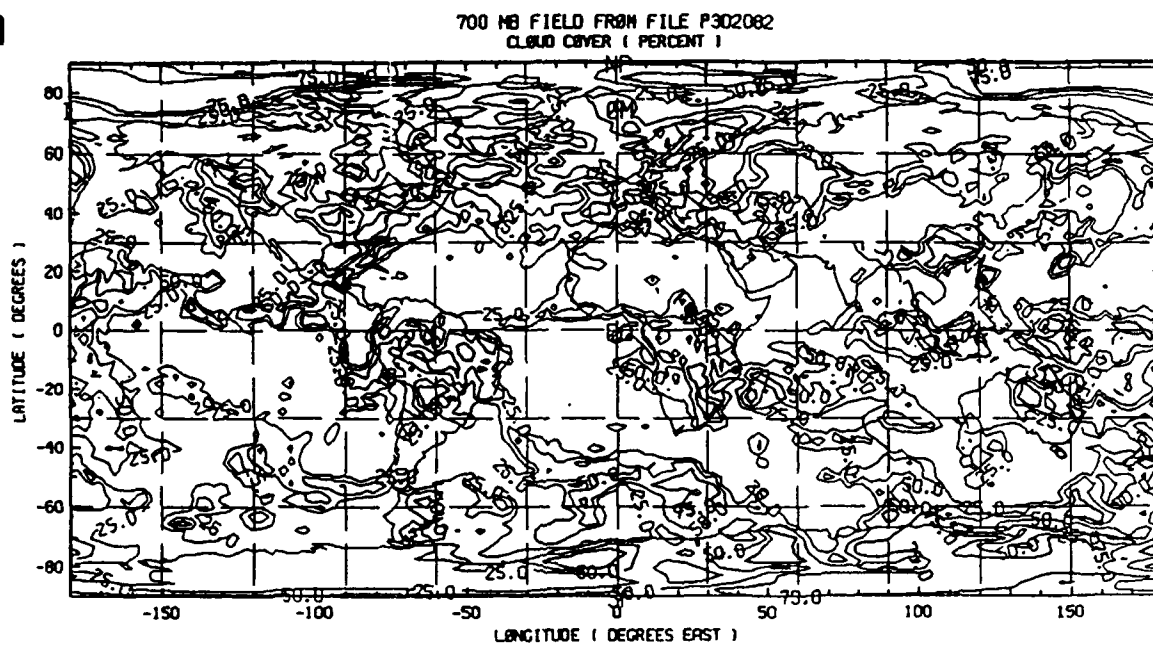


Fig. 7: Continued.

a



b

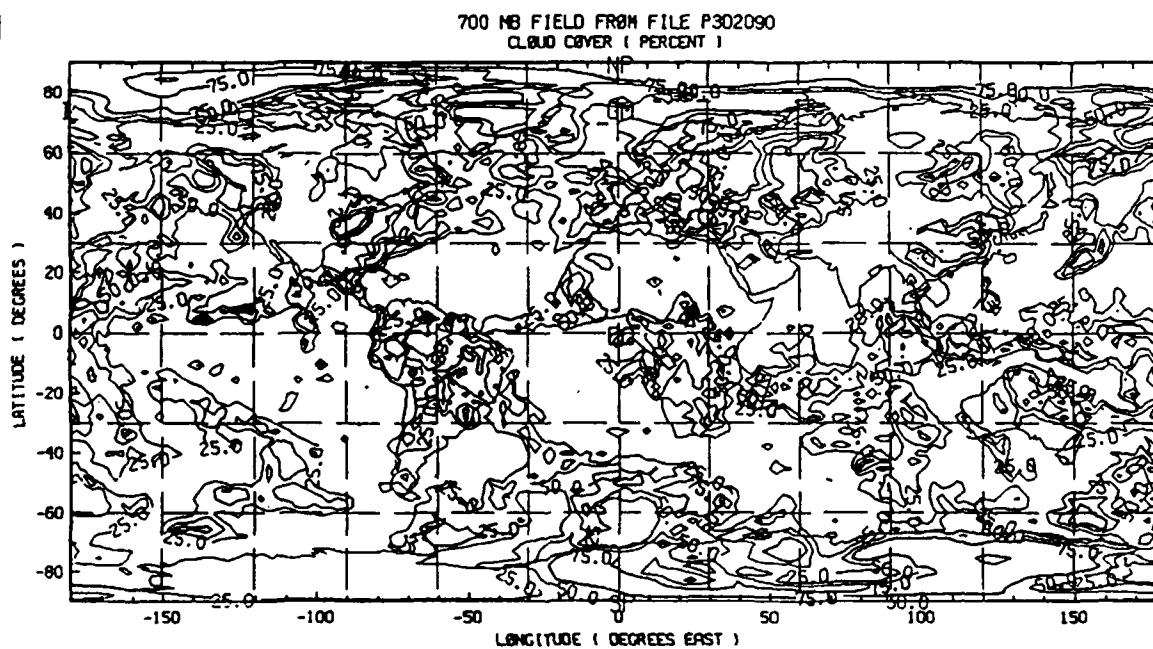


Fig. 8: Contour plots of 3DNEPH cloud data for February 8, 12Z (a), February 9, 00Z (b), and February 10, 12Z (c). Contours drawn at 25, 50, and 75%.

C

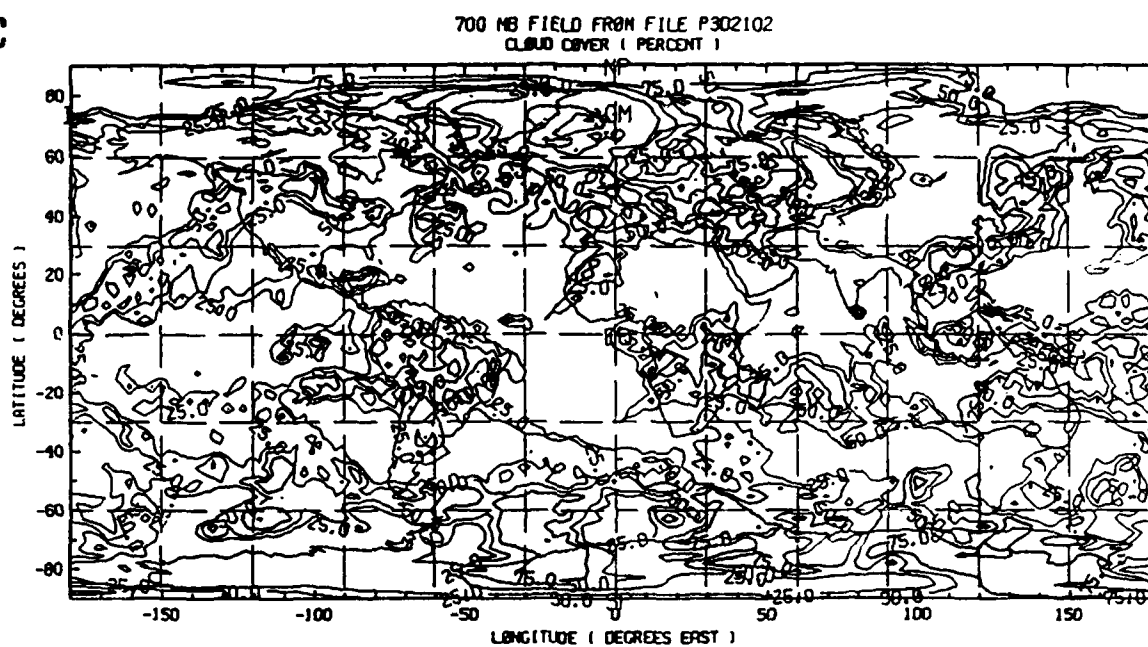


Fig. 8: Continued.

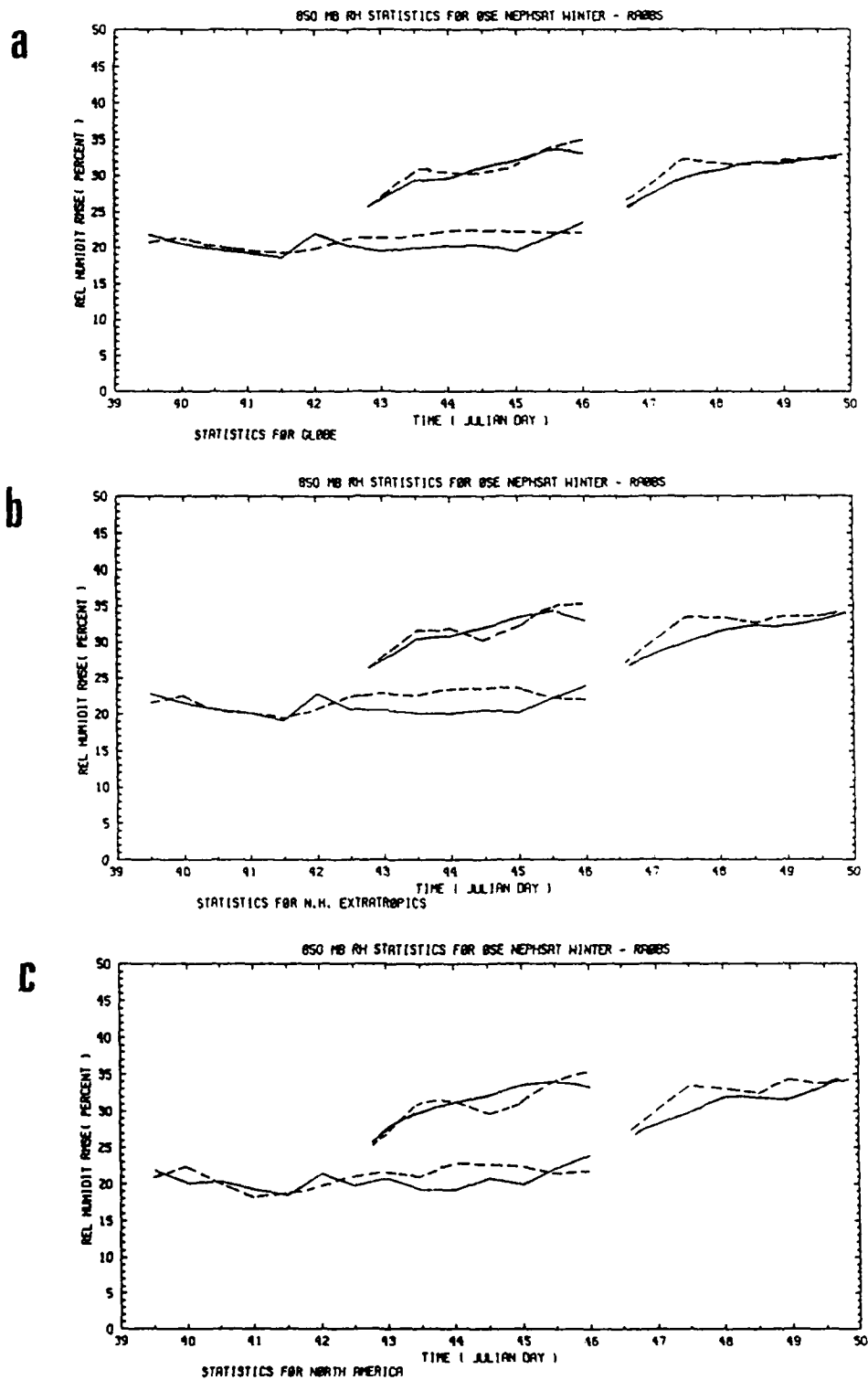
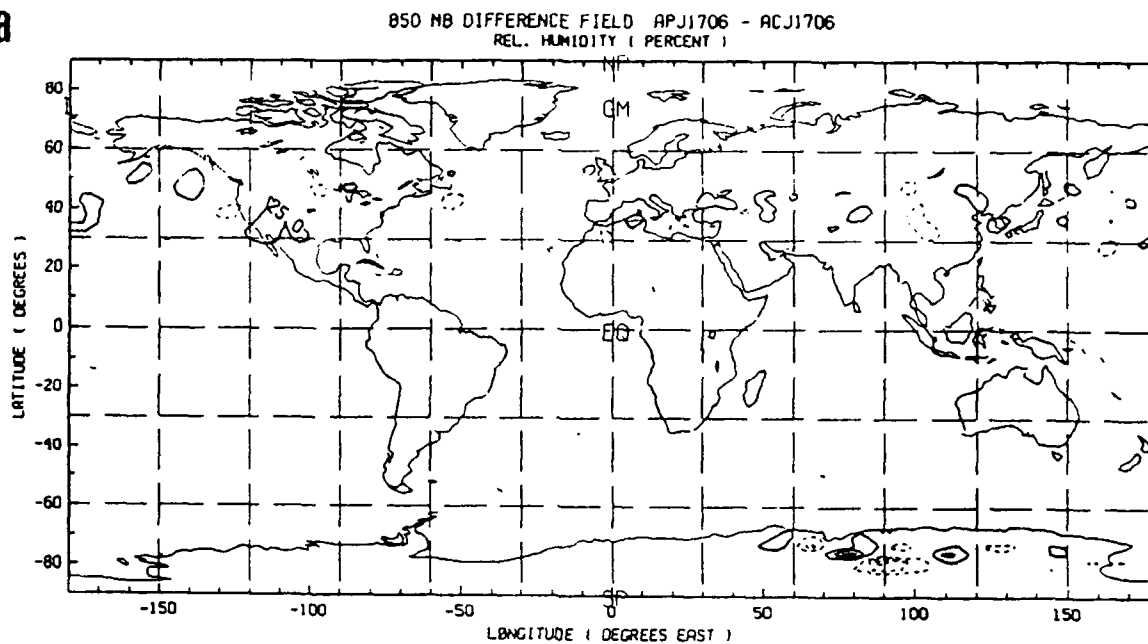


Fig. 9: Root mean square errors of the 850 hPa relative humidity analyses and forecasts, when compared with radiosonde data, for NEPHSAT (solid lines) and STATSAT (dashed lines). Statistics are shown for the globe (a), the Northern Hemisphere extratropics (b), and North America (c).

a



b

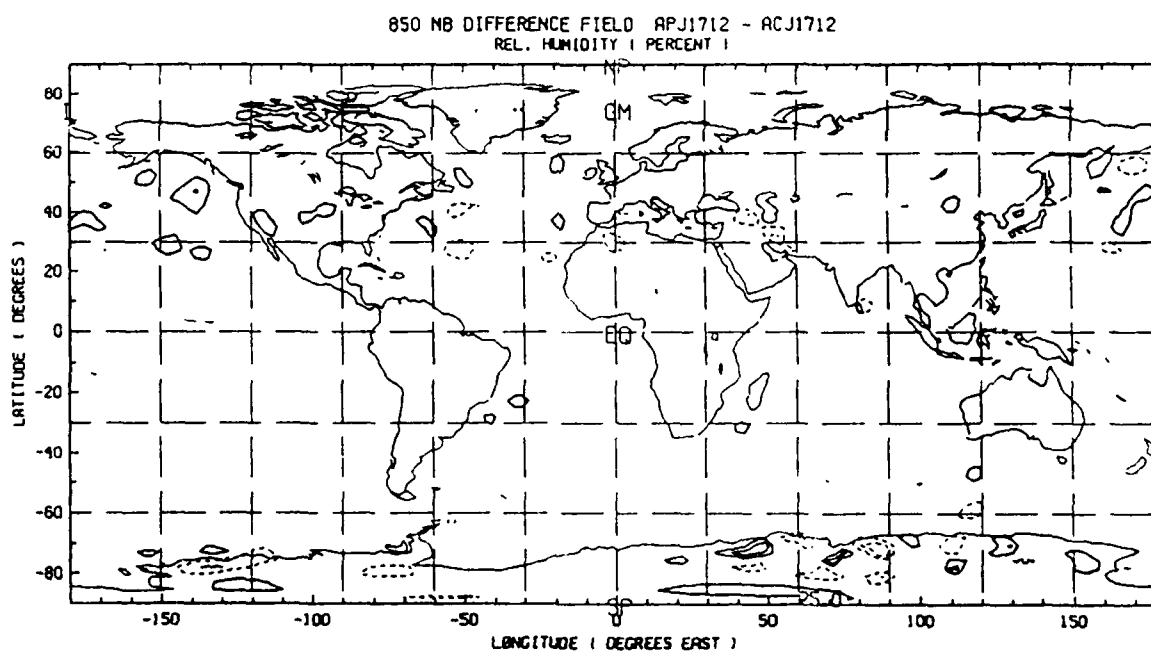
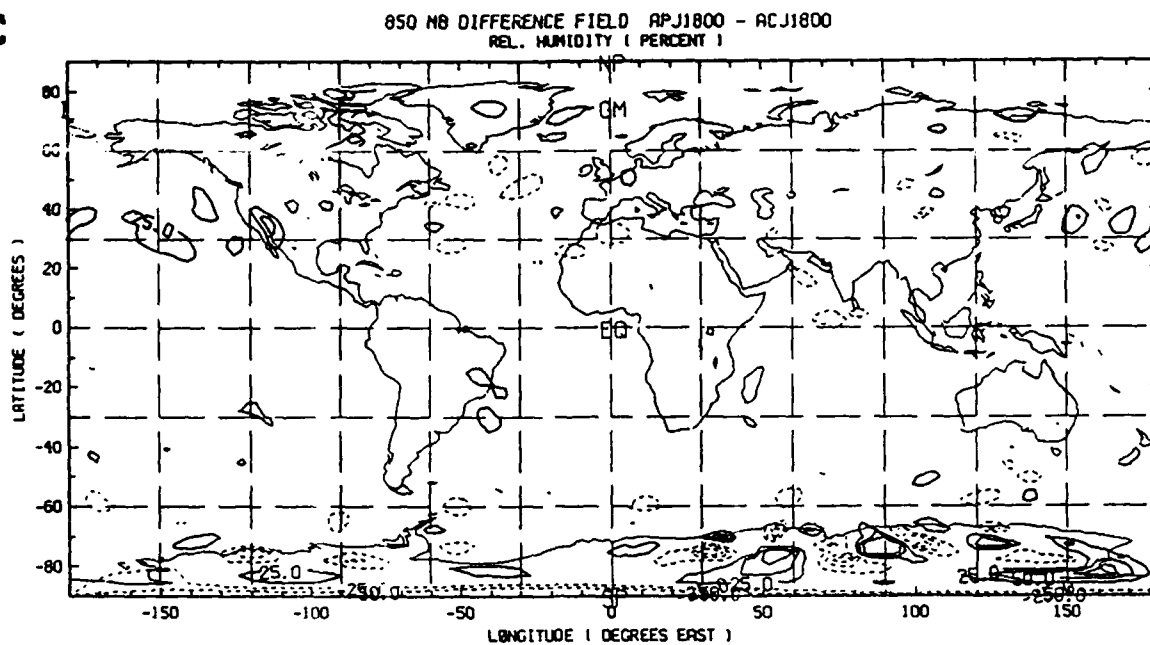


Fig. 10: As Fig. 8, but for June 17, 06Z (a), June 17, 12Z (b), June 18, 00Z (c), and June 19, 12Z (d).

c



d

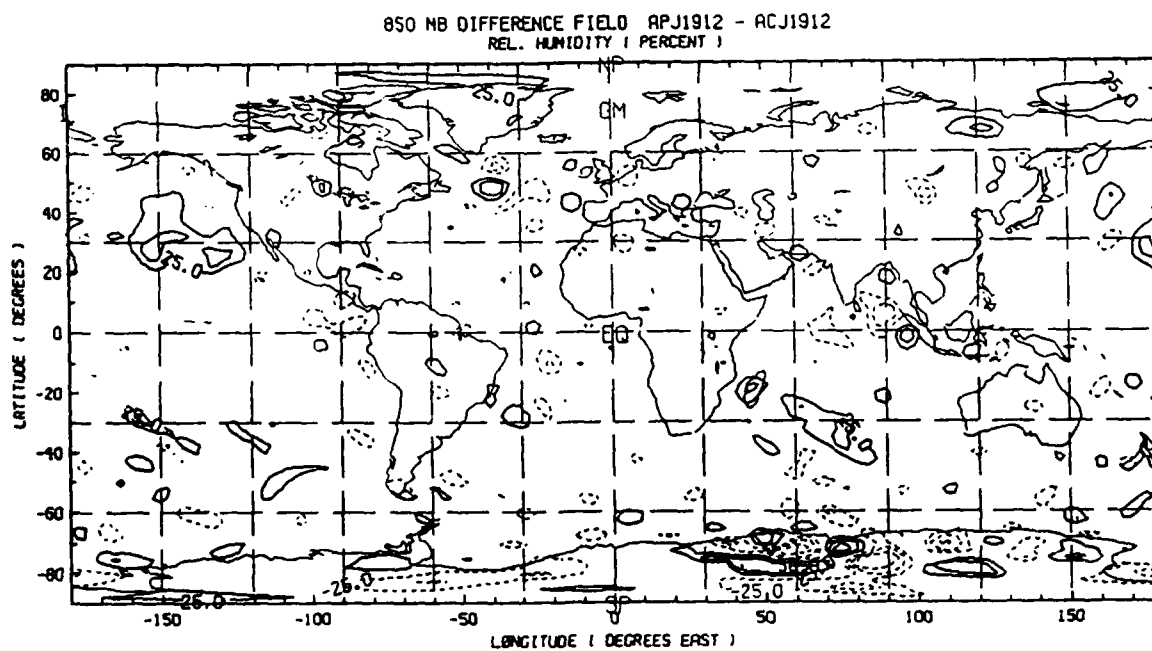
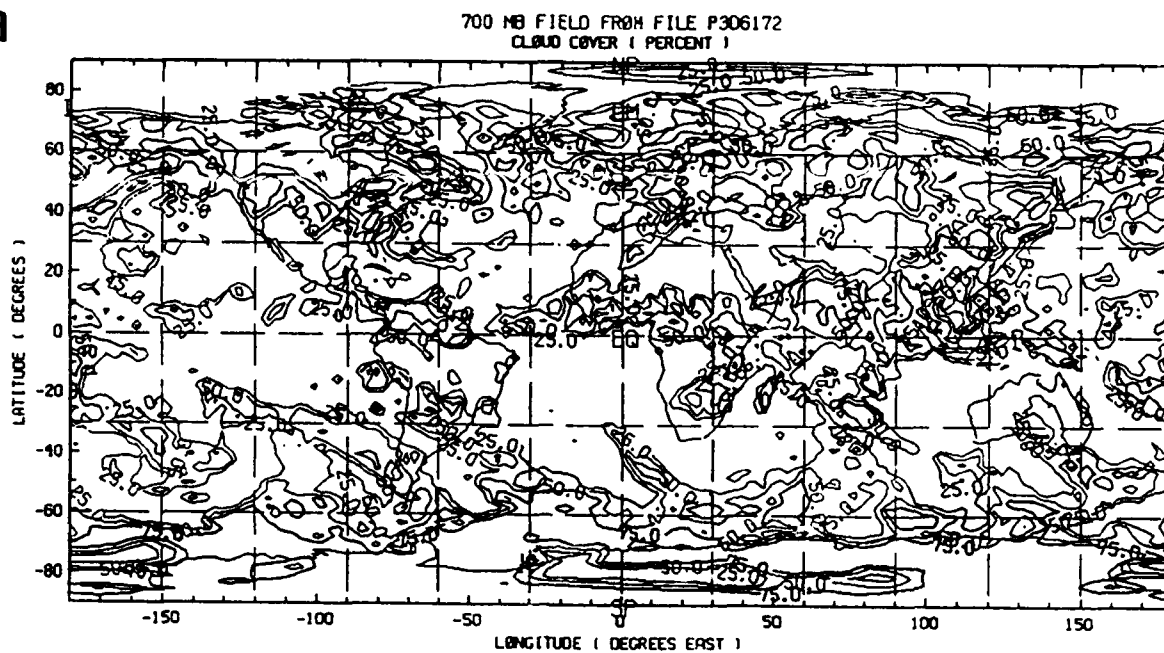


Fig. 10: Continued.

a



b

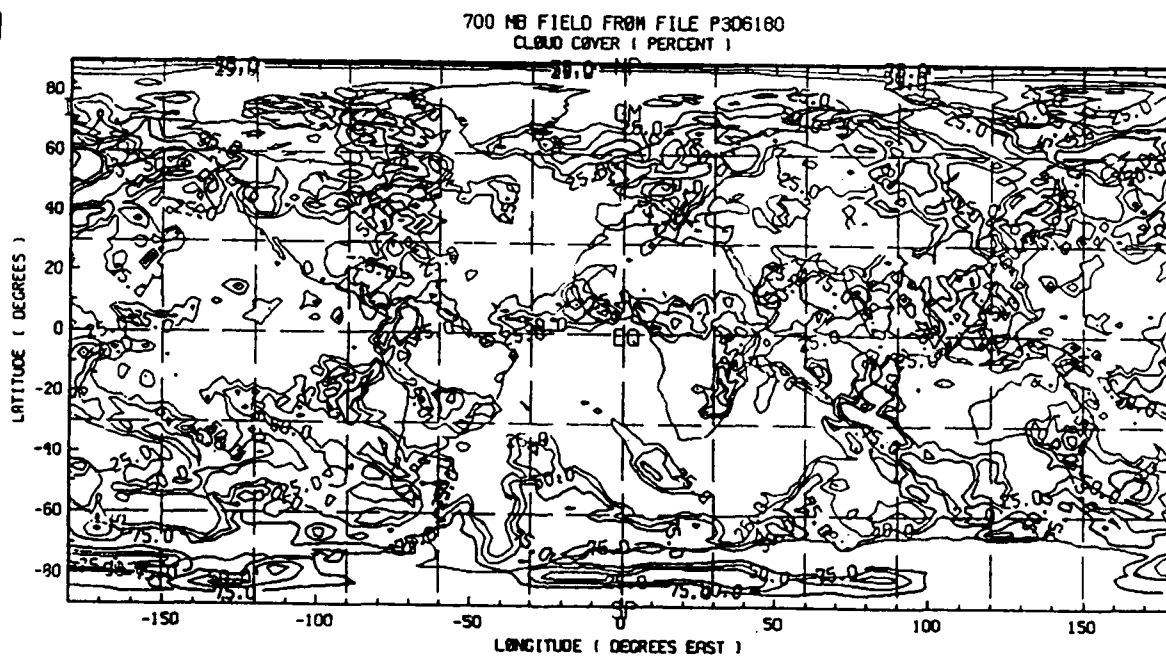


Fig. 11: As Fig. 9, but for June 17, 12Z (a), June 18, 00Z (b), and June 19, 12Z (c).

C

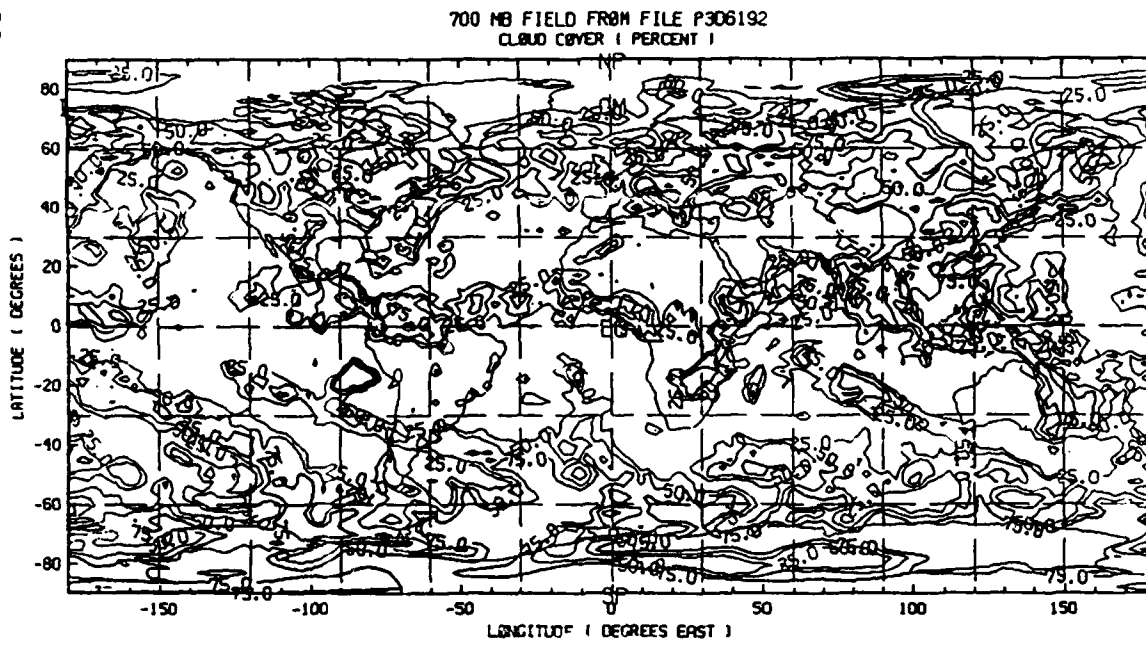


Fig. 11: Continued.